

# BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-67661

(43) 公開日 平成7年(1995)3月14日

(51) Int.Cl. <sup>8</sup>	識別記号	庁内整理番号	F I	技術表示箇所
C 1 2 N 15/09	Z N A			
A 6 1 K 38/00	A A R			
C 0 7 K 14/47		8318-4H		
		9050-4B	C 1 2 N 15/ 00	Z N A A
			A 6 1 K 37/ 02	A A R
審査請求 未請求 請求項の数22 O L (全 51 頁) 最終頁に続く				

(21) 出願番号	特願平6-36026	(71) 出願人	392015468 ザ・ジェネラル・ホスピタル・コーポレイ ション THE GENERAL HOSPITA L CORPORATION アメリカ合衆国02114マサチューセッツ州 ボストン、フルート・ストリート (番地 の表示なし)
(22) 出願日	平成6年(1994)3月7日	(72) 発明者	マーシー・イー・マクドナルド アメリカ合衆国02173マサチューセッツ州 レキシントン、ウォルサム・ストリート 462番
(31) 優先権主張番号	0 2 7 4 9 8	(74) 代理人	弁理士 青山 葆 (外1名)
(32) 優先日	1993年3月5日		最終頁に続く
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	0 8 5 0 0 0		
(32) 優先日	1993年7月1日		
(33) 優先権主張国	米国 (US)		

(54) 【発明の名称】 ハンチンチンDNA、そのタンパク質及びその用途

(57) 【要約】

【目的】 ハンチントン舞蹈病の診断及び処置を目的とする。

【構成】 ハンチンチンタンパク質をコードする新規な遺伝子ハンチンチン、それを発現できる組換えベクター及び宿主、並びにハンチンチン舞蹈病を診断し処置する方法に関する。

## 【特許請求の範囲】

【請求項1】 ハンチンチンタンパク質をコードする核酸を含有する単離された核酸。

【請求項2】 該ハンチンチンタンパク質が配列番号6に示すアミノ酸配列を有している請求項1に記載の核酸。

【請求項3】 該核酸が配列番号5に示すアミノ酸をコードするDNA配列を有している請求項2に記載の核酸。

【請求項4】 請求項1に記載の核酸又は少なくとも15個のその隣接ヌクレオチドを含有する、試料中のハンチンチンの存在を確認するための核酸プローブ。

【請求項5】 該プローブが配列番号5に示す核酸配列、又は少なくとも15個のその隣接ヌクレオチドを有している請求項4に記載の核酸プローブ。

【請求項6】 該プローブが配列番号6に示すアミノ酸配列又は少なくとも5個のその隣接アミノ酸をコードしている請求項4に記載の核酸プローブ。

【請求項7】 ハンチンチンタンパク質をコードする核酸が転写及び／又は翻訳発現シグナルと作動可能に連結している請求項1に記載の核酸。

【請求項8】 請求項7に記載の核酸を含有するベクター。

【請求項9】 請求項8に記載のベクターによって形質転換されている宿主細胞。

【請求項10】 配列番号6に示すアミノ酸配列を有するハンチンチンタンパク質を含有する細胞不含の組成物。

【請求項11】 ハンチンチンタンパク質又は少なくとも5個のその隣接アミノ酸に相当するアミノ酸配列を含有する実質的に純粋なタンパク質。

【請求項12】 配列番号6に示すアミノ酸配列を有する請求項11に記載の実質的に純粋なタンパク質。

【請求項13】 請求項11に記載のタンパク質又は少なくとも5個のその隣接アミノ酸をコードするRNA配列と相補的な配列と作動可能に連結されている、細胞内にて機能する転写領域を含有する組換え核酸分子。

【請求項14】 ハンチンチン抗体。

【請求項15】 請求項14に記載の抗体を産生するハイブリドーマ。

【請求項16】 患者におけるハンチントン舞蹈病の発現の存在又はその疾病素質を診断するための方法であって、

(a) 該患者から試料を採取し、  
(b) 該試料中のハンチンチン(CAG)<sub>n</sub>領域を検出することによって、該試料中のハンチンチン核酸の特性を評価し、

(c) 工程(b)の特性を、ハンチントン舞蹈病の疑いのない個体から入手した同様の分析値と比較し、

(d) これらのハンチンチン(CAG)<sub>n</sub>領域の特性が異

なる場合に、患者のハンチントン舞蹈病の発現の存在又はその疾病素質を診断することを特徴とする方法。

【請求項17】 ハンチントン舞蹈病患者の予後を確認するための方法であって、

(a) 該患者から試料を採取し、

(b) 該バイオプシー試料中のハンチンチン(CAG)<sub>n</sub>領域を含有するハンチンチン核酸の特性をサザンブロット分析又はポリメラーゼ鎖反応分析によって評価し、患者の予後を確認することを特徴とする方法。

【請求項18】 請求項1に記載の核酸を患者の細胞に付与することを特徴とする、そのような処置を必要としている患者のハンチントン舞蹈病を処置する方法。

【請求項19】 請求項13に記載の核酸分子を患者の細胞に付与することを特徴とする、そのような処置を必要としている患者のハンチントン舞蹈病を処置する方法。

【請求項20】 請求項11に記載のタンパク質に対するアンタゴニストを患者の細胞に付与することを特徴とする、そのような処置を必要としている患者のハンチントン舞蹈病を処置する方法。

【請求項21】 請求項11に記載のタンパク質に結合する化合物を患者の細胞に付与することを特徴とする、そのような処置を必要としている患者のハンチントン舞蹈病を処置する方法。

【請求項22】 該化合物がタンパク質である請求項21に記載の方法。

## 【発明の詳細な説明】

【0001】本発明の開発の際に行われた研究は米国政府の基金を利用している。米国政府は本発明について特定の権利を有している。

## 関連出願の相互参照

本出願は1993年3月5日に出願された米国特許出願第08/027,498号の一部継続出願である(この出願内容は引用によって本明細書に包含される)。

## 【0002】発明の分野

本発明は遺伝病の検出及び処置の分野に属する。詳細には、本発明はハンチンチン遺伝子(IT15遺伝子とも呼ばれる)、この遺伝子によってコードされているハンチンチンタンパク質、並びに(1)ハンチントン舞蹈病の発現の疾病素質を検出するため、(2)ハンチントン舞蹈病の診断のため、(3)ハンチントン舞蹈病の処置のため、及び(4)このような処置の処置経過のモニターのため、の検定におけるこの遺伝子及びタンパク質の用途に関する。

## 【0003】

【従来技術】ハンチントン舞蹈病(HD)は、運動障害、認識喪失及び精神医学症状発現を特徴とする進行性の神経変性障害である[Martin及びGusella, N R R L B-Engl.J.Med. 315:1267-1276(1986)]。これは常染色体の優性態様によって受け継がれ、欧州起源の殆どの

集団のおよそ1/10,000個体数が罹患している [Harper, P.S.ら, in Huntington's disease, W.B.Saunders, フィラデルフィア, 1991]。HDの顕著な特徴は、通常は人生の40年から50年にわたる油断のない潜行性の症状を有し、死亡するまでの10から20年にわたり徐々に悪化していく特有の舞踏運動障害である。時には、HDは硬直及びより迅速な経過をたどるより重篤な症状を伴って通常は現れる若年期に発現される場合がある。HDの若年期症状は疾患の対立形質の父系伝達の優位に関連している。HDの神経病因は、脳の尾部及び被殻領域にて最も重篤なニューロンの選択的な喪失を伴う特有のパターンをも呈する。HDにおけるニューロン死の生化学的基礎はこれまで説明されておらず、結局、この破壊的な障害の発現及び進行を遅らせ、又は予防するうえで有効な処置は存在しない。

【0004】HDを生起する遺伝欠損は、ヒトの多形DNAマーカーを使用する連鎖(リンケージ)分析の最初の成功例の1つとして、1983年に染色体4であると帰属された [Gusellaら, Nature 306:234-238(1983)]。この時から、本発明者らは、その位置についての累積的研究に基づき、HD遺伝子の単離及び特性化のための配置クローニング手法(location cloning approach)を探求した [Gusella, FASEB J. 3:2036-2041(1989); Gusella, Adv. Hum. Genet. 20:125-151(1991)]。他の研究の中には、多くの手法によってこの領域中に新たな遺伝子マーカーを創製し [Pohlら, Nucleic Acids Res. 16:9158-9198(1988); Whaleyら, Somat. Cell. Mol. Genet. 17:83-91(1991); MacDonaldら, J. Clin. Inv. 84:1013-1016(1989)]、関連領域の遺伝学 [MacDonaldら, Neuron 3:183-190(1989); Allittoら, Genomics 9:104-112(1991)] 及び身体地図 [Batesら, Nature Genet. 1:180-187(1992); Doucette-Stammら, Somat. Cell Mol. Genet. 17:471-480(1991); Altherrら, Genomics 13:1040-1046(1992)] を確立し、YACクローンにおいてHD染色体の4pテロメア(末端小粒)をクローニングし [Batesら, Am. J. Hum. Genet. 46:762-775(1990); Youngmanら, Genomics 14:350-356(1992)]、YAC(酵母人工染色体) [Batesら, Nature Genet. 1:180-187(1992)] 及び候補領域の(全配列を一緒に形成する一連の重複クローン)コスミド [Baxendaleら, 調製物にて] 整列群(contigs)を確立し、さらにこの領域由来の多くの候補遺伝子进行分析し特性化 [Thompsonら, Genomics 11:1133-1142(1991); Taylorら, Nature Genet. 2:223-227(1992); Ambroseら, Hum. Mol. Genet. 1:697-703(1992)] するものがある。HD血縁者における組換え事象の分析によって、最もHD遺伝子の位置であるらしい4p16.3内のD4S10及びD4S98間の2.2Mbの候補領域が同定された [MacDonaldら, Neuron 3:183-190(1989); Batesら, Am. J. Hum. Genet. 49:7-16(1991); Snellら, Am. J. Hum. Genet. 51:375-362(1992)]。HDと4p16.3内の

DNAマーカーとの間の連鎖の平衡異常 [Snellら, J. Med. Genet. 26:673-675(1989); Theilmanら, J. Med. Genet. 26:676-681(1989)] の調査により、多重突然変異が起こってその障害の原因となっていることが示された

[MacDonaldら, Am. J. Hum. Genet. 49:723-734(1991)]。しかし、多重対立遺伝子マーカーを使用する単相型の分析によって、少なくとも1/3のHD染色体が先祖伝来的に関連していることが示された [MacDonaldら, Nature Genet. 1:99-103(1992)]。これらのHD染色体が共有する単相型は、最も遺伝子欠損の部位らしいものとしてD4S180及びD4S182間の500kbセグメントを示した。

【0005】遺伝子転写体を完全にするため、この500kb領域をターゲットニングし、候補コード化配列を入手するための迅速な方法としてエキソン増幅を使用した [Bucklerら, Proc. Natl. Acad. Sci., U.S.A. 88:4005-4009(1991)]。この戦略により3つの遺伝子が既に同定されている: アデュシン(a-adducin)遺伝子(ADD A) [Taylorら, Nature Genet. 2:223-227(1992)]; このセグメントの遠位配置の推定の新規トランスポーター遺伝子(IT10C3); 及び中心部分における新規なGプロテイン結合レセプターキナーゼ遺伝子(IT11) [Ambroseら, Hum. Mol. Genet. 1:697-703(1992)]。しかし、これらの遺伝子が関連する欠陥としてHD座のようなものは見いだされなかった。

#### 【0006】発明の概要

本明細書にて「ハンチンチン(huntingtin)」又は「IT15」と呼ぶ、約210kbをまたぎ、従来記載されていない約348kDaのタンパク質をコードする巨大遺伝子が同定された。このハンチンチンの解読枠は、正常の種族に少なくとも17個の対立遺伝子を有する多形(CAG)<sub>n</sub>のトリヌクレオチドの反復(11から約34CAGのコピーで変動)を含有している。HD染色体上では、トリヌクレオチド反復の長さは実質的に増加しており、例えば約37から少なくとも73個のコピーであり、このことから症状発現(発症)の年令と見掛け上の相関性が認められ、最も長いセグメントが若年期HDの症例で同定されている。この反復の長さ不安定性は、脆X症候群及び筋緊張性ジストロフィー [Suthersら, J. Med. Genet. 29:761-765(1992)] における同様のトリヌクレオチド反復を想起させる。HD染色体上の不安定な伸長性のあるトリヌクレオチド反復が最も強固なその疾患との連鎖平衡異常の領域に存在することは、この変化がHDの優性表現型であり、ハンチンチンがHD遺伝子をコードしていることを示唆している。

【0007】本発明はハンチンチンタンパク質、そのタンパク質をコードするDNA及びRNA、並びにそれらの使用に関する。従って、本発明は第1の態様として、ハンチンチンタンパク質の精製された調製物に関する。別の態様として、本発明はハンチンチンをコードするD



NA又はRNAを含有する組換え構築物に関する。さらなる態様として、本発明はこのようなハンチンチンをコードする核酸を含有するベクターに関する。さらに、本発明はこのようなベクターによって形質転換されている宿主に関する。また、本発明はこのような組換え宿主からハンチンチンを生産する方法に関する。さらに、本発明はこのようなハンチンチンDNA、RNA及び／又はタンパク質を使用し、ハンチントン舞蹈病を診断する方法に関する。また、本発明はこのようなハンチンチンDNA、RNA及び／又はタンパク質を使用し、ハンチントン

10 舞蹈病を処置する方法に関する。  
【0008】さらには、本発明は、症状の又は前兆の患者の遺伝子治療の方法であって、正常範囲にある11-34コピーの(CAG)<sub>n</sub>を有する機能的なハンチンチン遺伝子をこのような処置を必要としている患者の所望の細胞に付与するに当たり、このような遺伝子によって付与されるハンチンチンタンパク質を該患者の細胞にハンチンチン機能を付与するに十分な時間及び量で発現できる態様で付与することを特徴とする方法に関する。また、本発明は、症状の又は前兆の患者の遺伝子治療の方法であって、機能的なハンチンチン・アンチセンス遺伝子をこのような処置を必要としている患者の所望の細胞に付与するに当たり、このような遺伝子によって付与されるハンチンチンアンチセンスRNAを該患者の細胞でのハンチンチンmRNA発現を阻害するに十分な時間及び量で発現できる態様で付与することを特徴とする方法に関する。

【0009】また、本発明は、症状の又は前兆の患者の遺伝子治療の方法であって、このような遺伝子を必要としている患者の細胞に機能的なハンチンチン遺伝子を付与することを特徴とする方法に関し、その1つの態様として該機能的なハンチンチン遺伝子は11-34コピーの(CAG)<sub>n</sub>サイズを含有している。さらに、本発明は、患者のハンチントン舞蹈病を診断し、又はハンチントン舞蹈病の発現についての素質(疾病素質)を診断する方法であって、該患者の、特に該患者の影響下組織におけるハンチントン遺伝子に存在する(CAG)<sub>n</sub>反復の数を測定することを特徴とする方法に関する。また、本発明は、患者のハンチントン舞蹈病を処置する方法であって、該患者の所望の細胞におけるハンチントン遺伝子のハンチントン(CAG)<sub>n</sub>反復の数を減少させることを特徴とする方法に関する。

#### 【0010】図面の説明

図1: HD候補領域の長い範囲の制限地図である。4p16.3の部分的な長い範囲の制限地図を示している[Linら, Somat. Cell Mol. Genet. 17:481-488(1991)から改作]。組換え事象によって決定したHD候補領域をD4S10及びD4S98間の斜線ラインで示している。連鎖の平衡異常単相型分析に基づく欠損部位として関連するHD候補領域の部分[MacDonaldら, Nature Ge

net. 1:99-103(1992)]は黒塗り部分として表している。この模式図の下方には、D4S180からD4S182の領域を伸ばし、コスミド整列群(平均40 kb/コスミド)を示している。ゲノム範囲及びハンチンチン(IT15)、IT11、IT10C3及びADDA遺伝子の転写方向(5'から3'への矢印)の知られている部分も示している。この地図の上方に示す座は、HD族に使用されている選択された多形マーカーを定義するものである。最大の平衡異常の領域における単相型の核を形成するD4S127及びD4S95の位置もコスミド整列群中に示している。制限部位はNotI(N)、MluI(M)及びNruI(R)として表している。完全な消化を示す部位は肉太文字で示し、不完全消化になる場合のある部位は細い文字で示している。「N」文字を囲む括弧は付加的な房のNotI部位の存在を示している。

【0011】図2: ハンチンチン(IT15)転写体のノーザンブロット分析である。正常(レーン1)及びHDホモ接合(レーン2及び3)リンバ芽球由来のRNAのノーザンブロットに対するIT15Aのハイブリダイゼーションの結果を示している。約11 kbの単一のRNAが3つすべての試料中で検出され、明らかな若干の変動はRNA濃度の差異に由来するものであった。HDホモ接合体は独立しており、それぞれ巨大アメリカ族(レーン2)及び巨大ベネズエラ族(レーン3)に由来するものであった。ベネズエラHD染色体はD4S127における(GT)<sub>n</sub>多形及びD4S95におけるVNTR及びTaqI RFLPsによって規定される「5 2 2」の4p16.3単相型を有している。アメリカホモ接合体は、HD染色体:「2 1 1 1」に見いだされる最も共通する4p16.3単相型を担持している[MacDonaldら, Nature Genet. 1:99-103(1992)]。

【0012】図3: IT15転写体を規定するcDNAクローンの模式図である。混成IT15配列の模式により、5つのcDNAを表している。細線は非翻訳領域に対応している。太線はコード化領域に対応しており、オープン読枠内の最初のMetコドンから翻訳が開始されられると思われる。星印は以下のエキソンのクローン5'から3'の位置を示している: DL83D3-8、DL83D3-1、DL228B6-3、DL228B6-5、DL228B6-13、DL69F7-3、DL178H4-6、DL118F5-U及びDL134B9-U4。この混成配列は以下のようにして誘導した。3'側22塩基から推定イニシエーターMet ATGまでの配列を、示したcDNAクローン及びエキソンから編集した。IT16Bの3'末端及びIT15Bの5'末端間には9塩基の配列が介在している。これらは最初の鎖cDNAのPCR増幅及びPCR産物の配列決定によるものである。混成配列の5'末端では、cDNAクローンIT16Cは(CAG)<sub>n</sub>の27塩基上流で終止する。しかし、IT16Cを同定した場合、新たな多形を生成さ

せる目的で(CAG)<sub>n</sub>の回りにゲノム配列を既に創製している。この配列はIT16C配列に適合しており、明らかなMet開始コドンを含めて337塩基上流に伸長している。

【0013】図4、図5、図6：ハンチンチン(IT15)(配列番号5及び配列番号6)の混成配列である。ハンチンチン(IT15)の混成DNA配列は配列番号5で示している。翻訳が長いオープン読枠の最初の枠内メチオニンから開始するとの仮定に基づき推定されるそのタンパク質産物(配列番号6)を、そのDNA

図7：(CAG)<sub>n</sub>反復のDNA配列分析である。パネル1、2及び3に示すDNA配列は、正常コスミドL191F1(1)、cDNA IT16C(2)、及びHDコスミドGUS72-2130にて検出される(CAG)<sub>n</sub>反復の変異を示している。パネル1及び3は、以下のプライマー(配列番号1)を使用するコスミド・サブクロンの直接配列決定によって作成した：

5' GGC GGG AGA CCG CCA TGG CG 3'

パネル2はpBSK11 T7プライマー(配列番号2)を使用して作成した：

5' AAT ACG ACT CAC TAT AG 3'

【0014】図8：若年期症状を示す幾つかの子供を用いたベネズエラHD同胞群の(CAG)<sub>n</sub>反復のPCR分析である。ベネズエラHD系統の同胞群のPCR分析結果を示している。罹患個体は黒記号で示している。子供は秘密保持のため三角で示している。AN1、AN2及びAN3は正常な染色体由来の対立遺伝子産物の位置を示している。AEはHD染色体由来のPCR産物の範囲を示している。バックグラウンドの一定バンドの強度は上記のPCR産物との比較のうえで有用であるが、それはPCR条件が若干相違して変動している。コスミドL191F1及びGUS72-2130由来のPCR産物はレーン12及び13に載せ、それぞれ18及び48のCAG反復を有している。

【0015】図9：同じHD単相型についてホモ接合の子供を用いたベネズエラHD同胞群の(CAG)<sub>n</sub>反復のPCR分析である。両親ともにHDに罹患しているベネズエラHD系統の同胞群のPCR分析結果を示している。子供は秘密保持のため三角で示し、ベネズエラ共同グループの研究者の盲検状態を保つため、HD診断情報は示していない。AN1及びAN2は正常な親染色体由来の対立遺伝子産物の位置を示している。AEはHD染色体由来のPCR産物の範囲を示している。コスミドL191F1及びGUS72-2130由来のPCR産物はレーン29及び30に載せ、それぞれ18及び48のCAG反復を有している。

【0016】図10：主要なHD単相型についてホモ接合の個体を用いたアメリカ族(アメリカファミリー)

の一員における(CAG)<sub>n</sub>反復のPCR分析である。主要なHD単相型が分離したアメリカ族の一員のPCR分析の結果。ANは正常な対立遺伝子の範囲を示し、AEはHD対立遺伝子の範囲を示している。レーン1、3、4、5、7及び8は関連HDの異種接合体由来のPCR産物を表している。レーン2は同じHD染色体についてホモ接合の族の一員由来のPCR産物を含有する。レーン6は正常な個体由来のPCR産物を含有する。系統の相関関係及び罹患状態は秘密保持のため示していない。コスミドL191F1及びGUS72-2130(レーン2に示す個体から誘導)をレーン9及び10に載せ、それぞれ18及び48のCAG反復を有している。

【0017】図11及び図12：HDを引き起こすと考えられる新たな突然変異による2つの族(ファミリー)の(CAG)<sub>n</sub>反復のPCR分析である。散发性のHD症例が新たな推定突然変異を表す2つの族のPCR分析結果を示している。各系統の個体を世代(ローマ数字)及び系統順序に基づき番号付けする。三角を使用し秘密を保持した。黒塗り記号は症状個体を示す。系統が分離している種々の染色体は4p16.3の多形マーカーを用いる膨大なタイピング(型別)によって区別され、ゲルのレーン上方に示す任意の数字に帰属された。星印を付けた染色体(図11の3、図12の1)は推定のHD染色体である。ANは正常な対立遺伝子の範囲を規定しており、AEは罹患個体及び同じ染色体を有する非罹患の親類に存在する対立遺伝子の範囲を規定している。

【0018】図13：対照染色体及びHD染色体上の(CAG)<sub>n</sub>反復単位の数比較である。150の独立した族由来の425HD染色体、及び545の対照染色体上にて観察される(CAG)<sub>n</sub>反復単位の数についての頻度分布を示している。

【0019】図14、図15：母親及び父親から伝達されるHD染色体上の(CAG)<sub>n</sub>反復単位の数比較である。母親(図14)及び父親(図15)からそれぞれ伝達されることが知られている図13由来の134及び161HD染色体に関する頻度分布を示している。2つの分布はT検定に基づき有意に相違している( $t_{133,159} = 5.34$ ,  $p < 0.0001$ )。

【0020】図16、図17、図18：3つの巨大な族由来のHD染色体上の(CAG)<sub>n</sub>反復単位の数異なるHD発端者(founders)と比較したものである。ベネズエラHD族(図16) [Gusella, J.F.ら, Nature 306: 234-238(1983); Wexler, N.S.ら, Nature 326:194-197(1987)]、Z族(図17)及びD族(図18) [Folstein, S.E.ら, Science 229:776-779(1985)]それぞれ由来の75、25及び35HD染色体に関する頻度分布を示している。ベネズエラの分布は図13の全体的HD染色体の分布と異なっていなかった( $t_{74,74} = 1.58$ ,  $p < 0.12$ )。Z族及びD族は共に全HD分布とは有意



に相違する分布を示した（それぞれ $t_{4,22}=6.73$ ,  $p<0.0001$ 及び $t_{4,30}=2.90$ ,  $p<0.004$ ）。

【0021】図19、図20： 両親及び対応する子供の(CAG) $n$ 反復長の相関関係を示す。母(図19)又は父(図20)のHD染色体上の反復長を対応する子供の反復長に対してプロットしている。総数25の母親伝達及び37の父親伝達がタイピングで入手できた。

【0022】図21： 精子及びリンパ芽球DNA由来のHD(CAG) $n$ 反復の増幅である。年令24-30のベネズエラHD系統の5人(1-5対)に関する精子(S)及びリンパ芽球(L)由来のDNAをHD(CAG) $n$ 反復のPCR増幅に使用した。各レーンにおける下方のバンドは正常な染色体由来である。

【0023】図22： 症状発現の年令と反復単位の長さとの相関関係を示している。症状発現の年令は、診断された234人のHD遺伝子キャリアーについて確立させ、それを、対応するリンパ芽球ラインのHD及び正常染色体の両方で観察された反復長さに対してプロットした。

#### 【0024】発明の詳しい説明

以下の説明では、分子遺伝学及び生物学における当業者に知られている種々の方法を引用している。このような引用した既知の方法を説明する文献及び他のものは、本明細書での引用によって完全に記載されているものとして本明細書に包含される。

【0025】本明細書にて説明しているIT15は、染色体4マーカーD4S180及びD4S182間の500kbセグメントの隣接部分由来の遺伝子である。ハンチンチン遺伝子は約210kb DNAをまたぎ、従来は記載されていなかった約348kDaのタンパク質をコードしている。ハンチンチンの解読枠は、正常ヒト集団において少なくとも17個の対立遺伝子を有する多形(CAG) $n$ トリヌクレオチド反復を含有し、この反復の数はこのような対立遺伝子にて11から約32のCAGコピーと変動する。これは、本明細書に示すように、ハンチントン舞蹈病において不安定な伸長した数のCAGトリヌクレオチド反復の存在に見舞われたヒト染色体の遺伝子であるので、ハンチンチン遺伝子内のCAG反復の数は37から少なくとも86コピー数の数にまで増大する。これらの結果は、ハンチンチン遺伝子が「ハンチンチン」と呼ばれるタンパク質をコードしているという結論及び、このようなハンチンチン遺伝子でのCAG反復の数の約37反復よりも大きな範囲での増大はハンチンチン舞蹈病の優性表現型に内包される変化であるという結論に基づいている。本明細書にて使用しているように、ハンチンチン遺伝子はハンチントン舞蹈病遺伝子とも呼ばれる。

【0026】遺伝子内のCAG反復の増幅を、ある遺伝子から転写されるこのような増幅CAG反復を含有する

mRNAを調節し、位置決めし、安定化し又は翻訳させる普通でない態様で行う以下の説明は、あらゆる遺伝子に適用可能であると考えられる。

#### 【0027】I. ハンチンチンDNAのクローニング及びハンチンチンタンパク質の発現

ハンチントン舞蹈病患者においてハンチンチンDNA及びそのタンパク質を改変遺伝子として同定する手法を以下に例示する。ハンチントン舞蹈病患者にてハンチンチン遺伝子の欠損を同定し、天然のヒトハンチンチン遺伝子を単離するための例示した方法及びその結果に加え、図4、図5及び図6に示す配列の情報は、線状又は環状組換え核酸構築物、例えばベクターに挿入し、それを使用して宿主細胞を形質転換する場合、本発明の方法によって天然のハンチンチンDNA及びハンチンチンタンパク質の有用な供給源であるハンチンチンDNA及びハンチンチンタンパク質のコピーを提供できる核酸及びタンパク質配列を提示するものである。このような方法は当業者に既知であるが、以下に概略説明する。

【0028】ハンチンチンのコード化配列を遺伝子操作する方法は所望のプロモーターの制御下の発現のために、このようなハンチンチンタンパク質をコードできる遺伝子配列のクローニングによって行う。このようなクローニング手法としては、ハンチンチンタンパク質をコードするDNA配列を構築するために当業者に知られている手法、例えば本明細書に記載しているハンチンチン配列を利用し、新たにハンチンチン遺伝子又はその遺伝子のCAG反復の数が変動しているその対立遺伝子を単離するポリメラーゼ鎖反応(PCR)手法、又は化学的方法を利用してヌクレオチド配列を構築するためのポリヌクレオチド合成などの手法を利用することができる。クローン化したハンチンチンDNAを発現させれば、ハンチンチンタンパク質が入手される。

【0029】本明細書にて使用している「遺伝子配列」なる用語は、DNA又はRNA、好ましくはDNAの核酸分子を意味するものである。ハンチンチンタンパク質をコードするDNAに作動可能に連結でき、その結果宿主細胞における発現及び維持を可能とする遺伝子配列は種々の供給源、例えば市販されている供給源、ゲノムDNA、cDNA、合成DNA、及びその組合わせ物から入手される。原核生物(細菌)宿主、真核生物宿主(植物、哺乳動物、特にヒト、昆虫、酵母及び特に培養細胞集団)などの宿主でハンチンチンタンパク質を発現させるには、遺伝子コードは普遍的であるので、本発明のハンチンチンアミノ酸配列をコードするあらゆるDNAが有用であると期待される。

【0030】ハンチンチン遺伝子を含有すると考えられるライブラリーからハンチンチンをコードする遺伝子を新たに選択することを望む場合は、ハンチンチン遺伝子又は発現されるハンチンチンタンパク質をコードする配列を特異的に選択する手段によってこのようなライブラ

リーをスクリーニングし、所望の遺伝子配列を同定すればよく、例えば(a) このタンパク質のDNA、例えば図4、図5及び図6に示しているようなもの、又はハイブリダイゼーション遺伝子を含むクロンを同定するに十分な長さである短くしたその機能的な誘導体に特異的な配列、を含む適当なハンチンチンDNAプローブ(群)を用いるハイブリダイゼーション(DNA:DNAハイブリダイゼーションの厳格な条件下)、又は(b) 問題のクロンにハイブリダイズする天然のハンチンチンmRNAをインビトロにて翻訳し、得られた翻訳産物をハンチンチンの生物活性の存在についてさらに特性化するハイブリダイゼーション選択翻訳分析、又は(c) ハンチンチンタンパク質を発現する宿主から翻訳されたハンチンチンタンパク質産物の免疫沈降、によって行えばよい。

【0031】ヒト対立遺伝子が図4、図5及び図6のそれとは同一の配列をコードしていない場合は、本明細書にて利用しているものと同じ手法、特にPCR手法によって、本明細書に開示した配列に基づくプライマーを用いて適当な遺伝子を増幅することによりハンチンチンDNAとして単離し同定することができる。染色体4上への遺伝子の繊細な位置決めは有用である多くの多形プローブは既知であり、入手可能である[例えば、ヒト及びマウスDNAプローブ及びライブラリーのATCC/NIH貯蔵カタログ、5編、1991、4-6頁を参照のこと]。例えば、有用なD4S10プローブはクロン名pTV20(ATCC57605及び57604);H5.52(ATCC61107及び61106)及びF5.53(ATCC61108)である。

【0032】ヒト染色体4に特異的なライブラリーは当業者に知られており、プローブを単離するためにATCCから入手可能であり[ヒト及びマウスDNAプローブ及びライブラリーのATCC/NIH貯蔵カタログ、5編、1991、72-73頁を参照のこと]、例えばLL04NS01及びLL04NS02(ATCC57719及びATCC57718)がこの目的に有用である。

【0033】本明細書に例示する正確なベクター構築物を利用する必要は必ずしもない。即ち当業者に知られている手法によって等価なベクターを構築すればよい。例えば、ハンチンチンDNAの配列は本明細書にて提示しており(図4、図5及び図6)、この配列により、ハンチンチン遺伝子の特異性が得られる。所望のプローブがこの配列、又はハンチンチン遺伝子の存在を陽性で指示するに十分なその一部を含むことが必須となるだけである。

【0034】ハンチンチンのゲノムDNAは天然に存在するイントロンを含むしても、又は含むしなくてもよい。さらに、このようなゲノムDNAは、その遺伝子配列の天然の5'プロモーター領域と一緒に、及び/又は天然のハンチンチン3'転写終止領域と共に入手しても

よい。

【0035】このようなハンチンチンゲノムDNAはまた、ハンチンチンmRNAの5'非翻訳領域をコードする遺伝子配列及び/又はハンチンチン3'非翻訳領域をコードする遺伝子配列と共に入手することもできる。宿主細胞がハンチンチンmRNA及びタンパク質の発現に関連する転写及び/又は翻訳調節シグナルを認識する限り、天然のハンチンチン遺伝子の5'及び/又は3'非翻訳領域、及び/又はハンチンチンmRNAの5'及び/又は3'非翻訳領域を保持させ、それらを転写及び翻訳の調節に用いることができる。

【0036】ゲノムDNAはあらゆる宿主細胞、特に染色体4を担持するヒト宿主細胞から当業者に周知の手段によって抽出し、精製することができる。ゲノムDNAは当業者に知られている手段、例えば物理的な裁断又は制限消化によって短くし、所望のハンチンチン遺伝子を染色体領域から単離することができ、またそうしなければその染色体領域は本発明の宿主にてハンチンチン遺伝子を利用するのに必要な情報以上のものが含有されている可能性がある。例えば、制限消化を利用すれば、所望の位置で完全長の配列を切断できる。あるいは、又はさらにDNA分子の3'末端から切断するヌクレアーゼを使用すれば、特定の配列を短い形態に消化することができ、次いでその所望の長さはポリメラーゼ鎖反応手法、ゲル電気泳動及びDNA配列決定によって同定し、精製される。このようなヌクレアーゼには例えば、エキソヌクレアーゼIII及びBcl31がある。他のヌクレアーゼは当業者には周知である。

【0037】あるいは、特定の宿主細胞集団がハンチンチンタンパク質を発現することが知られている場合は、当業者に知られているcDNA手法を利用し、このような集団に存在するハンチンチンmRNAのcDNAコピーを合成する。ハンチンチンタンパク質のアミノ酸配列をコードするゲノム又はcDNA核酸をベクターにクローニングするには、DNA調製物を適当なベクターに連結すればよい。ハンチンチンタンパク質をコードするDNA配列は、連結のための平滑末端又は粘着末端、適当な末端を付与する制限酵素消化、要すれば粘着末端の充填、望ましくない結合を回避するためのアルカリホスファターゼ処置、及び適当なリガーゼによる連結などの通常的手法によってDNAベクターに挿入することができる。このような操作を行うための手法は当業者に周知である。

【0038】ハンチンチンDNAをコードする配列及び作動可能に連結されたプロモーターを非複製、非挿入分子として受容真核生物細胞(好ましくはヒト宿主細胞)に導入する場合、そのコードされているハンチンチンタンパク質の発現は、導入配列の一時的(不安定)な発現によって起こすことができる。好ましくは、コード化配列は、自律的に複製できる閉じた環状又は線状分子など



のDNA分子に導入する。宿主染色体への導入を望む場合は、線状分子を用いるのが好ましい。染色体外要素としてハンチンチン遺伝子の安定な維持を望む場合は、所望の宿主において自律的に複製するのに適当なプラスミド要素を有する環状プラスミドの形態を使用するのが好ましい。

【0039】ハンチンチンタンパク質をコードする遺伝子及びそれと作動可能に連結された必須の調節要素を付与する所望の遺伝子構築物は形質転換、トランスフェクション、又は宿主細胞にその構築物を付与できるあらゆる方法によって望ましい宿主細胞に導入することができる。ハンチンチンDNAを受容した宿主細胞を検出するためのマーカー遺伝子はハンチンチンDNAと同じベクター上に、又はハンチンチンコード化配列構築物との宿主細胞への同時形質転換のための別の構築物上にあってもよい。ベクターの性質は宿主生物によって変動する。

【0040】適当な選択マーカーは宿主細胞によって変動する。例えば、マーカーは殺生物耐性、例えば抗生物質、銅などの重金属に対する耐性などを付与できるものである。具体的なプラスミド又はウイルスベクターを選択するうえで重要な因子には以下のものがある：ベクターを含有する受容細胞を認識し、その受容細胞をそのベクターを含有していない受容細胞から選択できる容易性；特定の宿主において望ましいベクターのコピーの数；及び別種の宿主細胞間でベクターを「往復運搬（シャトル）」できるのが望ましいか否か。

【0041】シャトルベクターの宿主として*S. cerevisiae*を使用するのが望ましい場合、好ましい*S. cerevisiae*酵母プラスミドには、2μサークルなど、又はその誘導体を含有するものがある。このようなプラスミドは当業者に周知であり、市販されている。ハンチンチン配列に特異的なオリゴヌクレオチドプローブを使用すれば、ハンチンチンに対するクローンを同定でき、またそのプローブは、本明細書に添付の図4、図5及び図6に示すタンパク質のアミノ酸配列の情報、又は同図に示すそのようなタンパク質又は関連タンパク質をコードするDNAの核酸配列の情報から、デ・ノボ(*de novo*)設計することができる。従って、ハンチンチンタンパク質に対して抗体を惹起させることができ、それを使用すれば、所望のクローン化タンパク質を発現する形質転換体における唯一のタンパク質決定基の存在を同定することができる。

【0042】DNAなどの核酸分子は、その核酸が転写調節情報を含有する発現制御配列を含有し、その配列がハンチンチンポリペプチドをコードするハンチンチンヌクレオチド配列と「作動可能に連結」されている場合に、ハンチンチンタンパク質を「発現することができる」と言われる。作動可能に連結とは、配列が調節配列（又は配列群）に結合されるに当たり、その態様がその

調節配列の影響下又は制御下に配列が発現されるように配置されている連結である。2つのDNA配列がコード化配列とそのコード化配列の5'末端に連結されたプロモーター領域配列である場合、プロモーター機能の誘導により所望のタンパク質をコードするmRNAが転写され、かつまたその2つのDNA配列の連結の性質が

（1）フレームシフト突然変異を誘導せず、（2）発現配列のタンパク質、アンチセンスRNAの発現能を妨害せず、又は（3）DNA鋳型の転写能を妨害しない場合は、それら2つの配列は作動可能に連結されている。従って、プロモーター領域がそのDNA配列を転写できる場合、そのプロモーター領域はDNA配列に作動可能に連結されていると言えるであろう。

【0043】遺伝子の発現に必要な調節領域の正確な性質は種又は細胞型に応じて変動し得るが、一般には転写及び翻訳の開始に関与するそれぞれ必須の5'非転写及び5'非翻訳（非コード化）配列、例えばTATAボックス、キャッピング配列、CAAT配列などがあり、プロモーター配列に必要なそれらの要素は本発明のプロモーターから提供される。このような転写制御配列には要すれば、エンハンサー配列又は上流のアクチベーター配列なども含有させることができる。本発明のベクターはさらに、作動可能に連結される他の調節因子、例えば1つ又はそれ以上の宿主細胞にてベクターを維持させるための抗生物質耐性又は複製起点を付与するDNA要素を含有することができる。

【0044】別の態様では、原核生物細胞又は酵母*S. cerevisiae*細胞にて本発明のベクターが維持できるために特に、受容宿主にて自律的に複製できるプラスミド又はウイルスベクターに導入配列を組込む。広範なベクターがこの目的で使用できる。バシラス(*Bacillus*)宿主では、所望のDNAの組込みが必要な場合がある。ヒト細胞などの真核生物宿主にてタンパク質を発現するには、このような宿主内で機能的な調節領域を使用することが必要である。宿主の性質に応じて、広範な転写及び翻訳調節配列が使用できる。好ましくは、これらの調節シグナルを、天然の状態で、特定のヒト組織タイプなどの特定のヒト細胞における発現レベルを高くできる特定の遺伝子と関連させる。真核生物では、転写が翻訳と連結していない場合、クローン化配列が開始メチオニン

（AUG）を含有しているか否かに応じて、このような制御領域は開始メチオニン（AUG）コドンが付与するか、又は付与しないものであってよい。このような領域は一般に、宿主細胞にてRNA合成を開始させるに見合うプロモーター領域が包含される。

【0045】望ましくは、ハンチンチンタンパク質をコードする配列の3'側の非転写及び／又は非翻訳領域は上記のクローニング方法によって入手することができる。天然のヒトハンチンチン遺伝子の3'非転写領域はその転写終止調節配列要素のため、又は真核生物細胞で



のポリアデニル化を行う要素のために保持させてもよい。天然の発現制御配列シグナルが宿主細胞にて十分に機能しない場合は、宿主細胞で機能する配列を置き換えることができる。

【0046】最初のタンパク質又は小さなペプチドの部分的なコード化配列（普通はそのアミノ末端）及びハンチンチンタンパク質のカルボキシ末端の別のコード化配列（部分又は全配列）を含有する融合産物を構築するのが望ましい場合がある。最初のタンパク質のコード化配列は例えば、ハンチンチンタンパク質を宿主細胞から分泌させるためのシグナル配列として機能するものであればよい。このような最初のタンパク質は多細胞生物の1つ細胞タイプにて作成され、それが同じ生物の別の細胞タイプに供給された場合、ハンチンチンタンパク質の組織ターゲティング又は局在化をも行うことができる。このような融合タンパク質の配列は所望のペプチド配列が以後の除去に反応するように、特定のプロテアーゼ部位を伴って、又は伴わずに設計することができる。

【0047】発現されたハンチンチンタンパク質は従来の条件、例えば抽出、沈降、クロマトグラフィー、アフィニティークロマトグラフィー、電気泳動などに従って宿主の培地から単離し精製することができる。例えば、抗ハンチンチン抗体を用いるアフィニティークロマトグラフィーを使用できる。図3に示すアミノ酸配列を有するタンパク質を作成でき、又はその配列の短鎖ペプチドを作成でき、それを使用して、当業者に周知の方法により抗体を惹起させる。これらの抗体は、ハンチンチンタンパク質を所望の供給源からアフィニティー精製し、又は定量化するために使用できる。ハンチンチンタンパク質を宿主細胞の細胞内領域から抽出する必要がある場合は、遠心によって、又は適当な緩衝液を用いて宿主細胞を採取し、細胞溶解し、そしてカラムクロマトグラフィー、例えばDEAE-セルロース、ホスホセルロース、ポリリボシチジル酸-アガロース、ヒドロキシアパタイト、又は電気泳動又は免疫沈降によってタンパク質を採取すればよい。

#### 【0048】11. 診断及び処置目的のためのハンチンチンの使用

以下の議論はヒト患者に具体的に焦点をあてているが、その教示はハンチンチンを発現し、ハンチンチンの改組、特にCAG反復コピー数の増幅がハンチンチン遺伝子（構造もしくは機能）又はハンチンチンタンパク質（構造もしくは機能もしくは発現）の欠損を導き、ハンチントン舞蹈病患者に認められるような臨床症状が認められる動物にも適用可能であると考えられる。また、本明細書に記載する方法はハンチントン舞蹈病が発現/発症するおそれのある、その状態が若年期あるいはそれよりも患者人生での高齢期のいずれで顕著になったかを問わず、そのような患者に適用できると考えられる。さらには、「患者」なる用語は、症状が現れていることを暗

示するものでなく、患者には、本発明の方法を使用して調査し又は処置するのが望ましい個体が包含されると考えられる。

【0049】本発明の診断及びスクリーニング方法は、家族歴に基づきハンチントン舞蹈病の発現の危険性が疑われる患者、又はハンチントン舞蹈病症状の存在を患者の症状の陰にある原因因子として診断し、又はその存在を排除することが望ましい患者にとって特に有用である。ハンチントン舞蹈病が診断されずに、患者の症状がハンチンチン遺伝子におけるCAG反復コピー数の変化に由来して起こったならば、本発明の方法はこのような症状の根底にある基礎のようなものを同定できる。

【0050】本発明のハンチンチン遺伝子をコードするDNA、特に正常ヒトハンチンチン遺伝子の配列を有するDNAを使用し、ハンチントン舞蹈病の発現の見込みをスクリーニングするのが必要な患者の前徴候スクリーニングが、本発明によって可能となった。本発明のスクリーニング方法によって、このような患者の異常なハンチンチン遺伝子の存在のための出生前診断などの前徴候診断が可能になり、従ってこのような個体がハンチントン舞蹈病又はその症状を発現しかならず、又は発現している見込みについて意見が述べられるようになったのである。このことは、例えばハンチントン舞蹈病の家族歴を有する個体由来のハンチンチン遺伝子のCAG反復の数が増大している改変ハンチンチン遺伝子対立遺伝子のキャリアーを同定するうえで特に有益である。患者のハンチンチン遺伝子のCAG反復の数を測定するため、そのような領域を増幅するためのPCRを使用し、又はDNAブロッティング手法を使用することは特に有用である。

【0051】例えばスクリーニング方法においては、組織試料をそのような個体から採取し、(1)「正常な」ヒトハンチンチン遺伝子の存在、特にこのような遺伝子の11-34CAGコピーの「正常」範囲の存在をスクリーニングする。ヒトハンチンチン遺伝子は例えば、本発明にて教示されるハンチンチン配列（又はその機能的断片）に対して調製したDNAプローブを用いるRFLP分析などの、「正常」DNA及び患者DNAにおける制限消化パターンの検出に基づいて特性化することができる。同様に、ハンチンチンmRNAは同様のプローブを用いて、ハンチントン舞蹈病発症の危険性のないヒト集団に見いだされるような正常ハンチンチンmRNAの(a)レベル及び/又は(b)サイズと比較し、特性化することができる。結論として、ハンチンチンタンパク質は例えば、免疫学的検定及び抗ハンチンチン抗体を用いるハンチンチンの生物検定によって、(a)検出され、及び/又は(b)定量することができるのである。ハンチンチンタンパク質を検定する場合、そのスピードの点から免疫学的検定が好ましい。ハンチンチンに対する抗体を作成する方法は当業者に周知である。

【0052】(1) 異常なハンチンチンDNAサイズのパターン、例えば異常なハンチンチンRFLP、及び／又は(2) 異常なハンチンチンmRNAのサイズ又はレベル、及び／又は(3) 異常なハンチンチンタンパク質レベルは、その患者がハンチントン舞踏病に関連する症状などのハンチンチン関連症状を発現し、又はそれを発現(発症)する危険にあることを示している。本発明のスクリーニング及び診断方法はプローブとして全ハンチンチンDNAコード化配列を使用する必要はない。むしろ、正常又は罹患個体由来のDNA調製物中のハンチンチン遺伝子の存在、そのような遺伝子の不存在、又はそのような遺伝子の改変した物理的性質(例えば、電気泳動の移動パターンの変化)を検出するに十分な核酸の断片又は長さを使用する必要があるのみである。

【0053】要すれば、胎児細胞を入手する既知の方法、例えば羊水穿刺、絨毛膜絨毛サンプリング(CVS)及び胎児撮影などの出生前診断を行うことができる。出生前の染色体分析は、正常ハンチンチン遺伝子を担持する染色体4の部分に異種接合状態で存在しているか否かを決定するのに使用でき、PCR増幅又はDNAプロットングが、ハンチンチン遺伝子に存在するCAGのサイズを算定するために利用される。ハンチンチンDNAは合成することができ、特にCAG反復領域は増幅でき、要すれば当業者に既知の手法によって放射性又は非放射性レポーターグループにより標識することができる[例えば、Eckstein, F.編, Oligonucleotides and Analogue: A Practical Approach, IRL Press at Oxford University Press, ニューヨーク, 1992; 及びKrucka, L.J.編, Nonisotopic DNA Probe Techniques, アカデミック・プレス, サンジエゴ(1992)を参照のこと]。

【0054】ハンチントン舞踏病の処置を必要とする患者におけるハンチントン舞踏病を処置する1つの方法では、機能的なハンチンチンDNAをこのような患者の細胞に提供し、好ましくは、本発明の方法を標的化するのが望まれる患者における多くの神経細胞の死亡を示すような徴候状態が現れる前に提供する。ハンチンチンDNAの置換は、このような遺伝子によって提供されるハンチンチンタンパク質をその患者を処置するに十分な時間及び量で発現できる態様かつそのような趣旨で提供する。細胞から欠失している遺伝子又はタンパク質が必要であるヒト患者にこのような態様で供給できるベクター系は数多く当業者に知られている。例えば、アデノウイルス又はレトロウイルス系、特に改変レトロウイルス系及び特に単純ヘルペスウイルス系を使用できる。このような方法は例えば、Breakefield, X.A.ら, The New Biologist 3:203-218(1991); Huang, Q.ら, Experimental Neurology 115:303-316(1992)、WO93/03743及びWO90/09441の教示から得られる[これらは引用によって完全に本明細書に包含される]。アンチセ

ンス戦略の方法は当業者に知られている[例えば、Anti sense Strategies, Baserga, R.ら, 編, Annals of the New York Academy of Sciences, 660巻, 1992]。

【0055】ハンチントン舞踏病の処置を必要とする患者のハンチントン舞踏病を処置する別の方法では、ハンチンチンアンチセンスRNAを転写する発現可能な配列をコードする遺伝子をこのような患者に提供し、好ましくは本発明の方法を標的化するのが望まれる患者における多くの神経細胞の死亡を示すような徴候状態が現れる前に提供する。ハンチンチンアンチセンスRNA遺伝子の置換は、このような遺伝子によって提供されるアンチセンスRNAをその患者を処置するに十分な時間及び量で、特にこのような患者の細胞に発現される異常なハンチンチンmRNAの翻訳を阻害できる量で発現できる態様かつそのような趣旨で提供する。上述のように、患者細胞にて改変している遺伝子又はタンパク質を必要とするヒト患者にこのような態様で供給できるベクター系は数多く当業者に知られている。例えば、アデノウイルス又はレトロウイルス系、特に改変レトロウイルス系及び特に単純ヘルペスウイルス系を使用できる。このような方法は例えば、Breakefield, X.A.ら, The New Biologist 3:203-218(1991); Huang, Q.ら, Experimental Neurology 115:303-316(1992)、WO93/03743及びWO90/09441の教示から得られる[これらは引用によって完全に本明細書に包含される]。

【0056】図4、図5及び図6に示すアミノ酸コード化配列などの機能的なハンチンチンタンパク質をコードするDNA配列を供給すれば、本発明の改変ハンチンチン遺伝子と効果的に置換され、ハンチントン舞踏病発症の危険性のないヒト集団に見いだされる11-34CAG反復と比較して増加しているハンチンチン遺伝子内のCAG反復の数、例えば37-86に由来するハンチンチン遺伝子発現に対する影響の結果である徴候を抑制し、及び／又は停止させ、及び／又は緩解させる。ハンチントン舞踏病は脳の尾状及び被殻領域で最も重篤である神経喪失が特徴であるので、本発明の処置方法は、細胞死に由来して多くの神経が喪失する前に置換ハンチンチン遺伝子をその疾患の経過初期に患者に投与すると最も有効である。この理由から、本発明の前徴候スクリーニング方法は本発明の方法による処置を必要とする個体を同定するうえで重要であり、このような処置はこのような個体が前徴候である際に行うのが好ましい。

【0057】ハンチントン舞踏病を処置する必要のある患者のハンチントン舞踏病を処置するさらなる方法では、このような患者の細胞に異常なハンチンチンタンパク質に対するアンタゴニストを投与する。本発明の方法をDNA-DNAプローブについて詳細に説明しているが、本発明のDNAと同じ配列情報を有するRNAも要すれば使用することができる。診断検定のためには、ハンチンチン抗体がハンチンチンタンパク質レベルの定量



及び評価に有用であり、特に免疫検定及び診断キットに有用である。

【0058】本発明は1つの態様として、ハンチンチンポリペプチド又はその結合断片に対する結合親和性を有する抗体に関する。好ましい態様では、そのポリペプチドは配列番号6に記載のアミノ酸配列を有しており、又はその突然変異体もしくはその種変異体、又は少なくとも7個のその隣接アミノ酸（好ましくは、少なくとも10、15、20又は30のその隣接アミノ酸）である。ハンチンチンを含む組織における改変ハンチンチンの発現を分析する（この分析に限定されるべきでない）などの方法に使用するには、ハンチンチンと特異的に結合するものを選択する。

【0059】本発明の抗体にはモノクローナル抗体及びポリクローナル抗体、並びにそれらの抗体の断片が包含される。その分子のイデオタイプを含有する抗体断片は既知の手法によって生成させることができる。例えば、このような断片にはF(ab')<sub>2</sub>断片、F(ab')断片、及びFab断片があるが、これらに限定されない。本発明にて特に興味深いのは、ヒトにて産生されるハンチンチン（又はその機能的誘導体）に対する抗体、又は組換え手法又は他の手法によって「擬人化（ヒト化）」されているもの（即ち、ヒトにおいて非免疫原性のもの）である。擬人化抗体は例えば、抗体の免疫原性部分を、対応する非免疫原性の部分と置き換えることによって調製することができる（即ち、キメラ抗体である）[Robinson, R.R.ら, 国際特許公開PCT/US 86/02269; Akira, K., 欧州特許出願184, 187号; Taniguchi, M., 欧州特許出願171, 496号; Morrison, S.L.ら, 欧州特許出願173, 494号; Neuberger, M.S.ら, PCT出願WO 86/01533; Cabilly, S.ら, 欧州特許出願125, 023号; Better, M.ら, Science 240:1041-1043(1988); Liu, A.Y.ら, Proc. Natl. Acad. Sci., U.S.A. 84:3439-3443(1987); Liu, A.Y.ら, J. Immunol. 139:3521-3526(1987); Sun, L.K.ら, Proc. Natl. Acad. Sci., U.S.A. 84:214-218(1987); Nishimura, Y.ら, Canc. Res. 47:999-1005(1987); Wood, C.R.ら, Nature 314:446-449(1985); Shawら, J. Natl. Cancer Inst. 80:1553-1559(1988)を参照のこと]。「擬人化」キメラ抗体のついての一般的概論はMorrison, S.L. [Science 229:1202-1207(1985)] 及びOi, V.T.ら, BioTechniques 4:214(1986)によって教示されている。適当な「擬人化」抗体はあるいは、CDR又はCEA置換 [Jones, P.T.ら, Nature 321:552-555(1986); Verhoevenら, Science 239:1534(1988); Beidler, C.B.ら, J. Immunol. 141:4053-4060(1988)] によって調製することができる。

【0060】別の態様では、本発明は上記のモノクローナル抗体又はその結合断片を産生するハイブリドーマに関する。ハイブリドーマは特異的なモノクローナル抗体を分泌できる不滅化セルラインである。一般に、モノク

ローナル抗体及びハイブリドーマを調製する手法は当業者に周知である [Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology," エリセビア・サイエンス出版, アムステルダム, ザ・ネザーランド(1984); St. Groth編, J. Immunol. Methods 35:1-21(1980)]。

【0061】抗体を産生することが知られている動物（マウス、ウサギなど）を選択ポリペプチドで免疫すればよい。免疫の方法は当業者に周知である。このような方法にはポリペプチドの皮下又は腹腔内注射がある。当業者であれば、免疫に使用するポリペプチドの量は免疫する動物、ポリペプチドの抗原性及び注射部位によって変動することは理解されよう。ポリペプチドはそのペプチド抗原性を増大させるためアジュバンド中にて修飾し、投与することができる。ポリペプチドの抗原性を増大させる方法は当業者に周知である。このような手法には、抗原と異種タンパク質（例えばグロブリン又はβ-ガラクトシダーゼ）とのカップリング、又は免疫の際にアジュバンドを包含させるなどがある。

【0062】モノクローナル抗体に関しては、免疫した動物から脾臓細胞を取り出し、骨髓腫細胞と融合させれば、モノクローナル抗体を産生するハイブリドーマ細胞を得ることができる。当業者に周知の多くの方法の1つを使用し、所望の特性を有する抗体を産生するハイブリドーマ細胞を同定できる。これにはELISA検定、ウェスタンブロット分析、又はラジオイムノアッセイを使用するハイブリドーマのスクリーニングなどがある

[Lutzら, Exp. Cell Res. 175:109-124(1988)]。所望の抗体を分泌するハイブリドーマをクローンし、当業者に知られている手法 [Campbell, Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, 前掲(1984)] によってそのクラス及びサブクラスを決定する。ポリクローナル抗体に関しては、抗体を含有する抗血清を免疫した動物から単離し、それを上記の手法の1つによって、所望の特異性を有する抗体の存在に関してスクリーニングする。

【0063】本発明の別の態様では、上記の抗体を検出可能に標識する。放射性同位体、親和性標識（例えば、ビオチン、アビジンなど）、酵素標識（例えば、西洋ワサビペルオキシダーゼ、アルカリホスファターゼなど）、蛍光性標識（例えば、FITC又はローダミンなど）、常磁性元素などを用いて、抗体は検出可能に標識することができる。このような標識化を行う手法は当業者に周知である [例えば、Sternbergerら, J. Histochem. Cytochem. 18:315(1970); Bayerら, Meth. Enzym. 62:308(1979); Enqvistら, Immunol. 109:129(1972); Coding, J. Immunol. Meth. 13:215(1976)を参照のこと]。本発明の標識抗体は、特異的なペプチドを発現する細胞又は組織を同定するため、インビトロ、インビボ及びin situ 検定に使用できる。

【0064】本発明のもう1つの態様では、上記の抗体を固体支持体に固定化する。このような固体支持体を例示すれば、ポリカーボネートなどのプラスチック、アガロース及びセファロースなどの複合炭水化物、アクリル樹脂、及びポリアクリルアミド及びラテックスビーズなどが挙げられる。このような固体支持体に抗体をカップリングする手法は当業者に周知である〔Weirら, "Handbook of Experimental Immunology" 4編, ブラックウェル・サイエンティフィック出版, オックスフォード, 英国, 10章(1986); Jacobyら, Meth.Enzym.34 アカデミック・プレス, N.Y.(1974)〕。本発明の固定化抗体はインビトロ、インビボ、及びin situ検定、並びに免疫クロマトグラフィーに使用することができる。

【0065】さらに、当業者ならば、現在利用可能な手法、及び抗体に関して先に説明した技法、方法及びキットを改作し、合理的に設計した抗ペプチドペプチドを創製するため、特異的なペプチド配列に結合できるペプチドを容易に作成できるであろう。例えば、Hurbyら, "Application of Synthetic Peptides: Antisense Peptides", In Synthetic Peptides, A User's Guide, W.H.Freeman, NY, 289-307頁(1992)、及びKaspczakら, Biochemistry 28:9230-8(1989)を参照のこと。抗ペプチドペプチドは2つの方法の一方によって作成することができる。まず、抗ペプチドペプチドを、ハンチンチンペプチド配列内に見いだされる塩基性アミノ酸残基を疎水性及び非帯電極性基を維持しつつ酸性残基と置換して作成すればよい。例えば、リジン、アルギニン及び／又はヒスチジン残基をアスパラギン酸又はグルタミン酸と置換し、グルタミン酸残基をリジン、アルギニン又はヒスチジンと置換する。本発明を実施する態様及び方法は以下の実施例を参照すれば、当業者ならばより完全に理解できるであろうが、これらの実施例はあらゆる意味において本発明の範囲又はそれを目的とする請求の範囲を限定するものではない。

#### 【0066】実施例

ハンチントン舞蹈病の原因となる遺伝子を4p16.3にてマッピングしたが、これは以前には同定されていなかった。本発明は、欠損の部位と思われる4p16.3の小さなセグメントに焦点をあてるため、連鎖平衡異常の単相型分析を使用した。新しい遺伝子であるハンチンチン(HTT)は標的領域のコスミド整列群からクローン化した「トラップ」エキソンを用いて単離されたが、これはHD染色体上にて伸長し不安定である多形トリヌクレオチド反復を含有している。広範な民族背景と4p16.3単相型とからなる調査した75の疾患ファミリーすべてのHD染色体上に、正常範囲である約11-約34コピーよりも長い(CAG)<sub>n</sub>反復が観察された。HD染色体上にて37から少なくとも86のコピーで変動する(CAG)<sub>n</sub>反復は、広く発現されるが既知の遺伝子とは関連しない約348 kDaの推定タンパク質の

コード化配列内に明らかに位置している。従って、ハンチントン舞蹈病の突然変異には、脆X症候群及び筋緊張性ジストロフィーにて記載されていると同様の不安定なDNAセグメントが関与しており、新規な4p16.3遺伝子の関連で機能し、優性な表現型を生じさせている。

【0067】次のプロトコール及び実験の詳細は、以下の実施例にて引用されている。

HDセルライン： 遺伝子連鎖及び平衡異常の研究に使用した種々の民族背景のHDファミリー由来のリンパ芽球セルライン〔Conneallyら, Genomics 5:304-308(1989); MacDonaldら, Nature Genet. 1:99-103(1992)〕はモレキュラー・ニューロジェネチクス大学、マサチューセッツ・ジェネラル・ホスピタルにおいて13年も前から樹立されている〔Anderson及びGusella, In Vitro 20:856-858(1984)〕。ベネズエラHD系統は、すべての罹患個体が共通する祖先(founder)からHD遺伝子が遺伝されている10,000人の広範な血縁者である〔Gusellaら, Nature 306:234-238(1983); Gusellaら, Science 225:1320-1326(1984); Wexlerら, Nature 326:194-197(1987)〕。

【0068】DNA/RNAプロットティング： 培養細胞からDNAを調製し、DNAプロットを作成し、記載されているようにハイブリダイズした〔Gusellaら, Proc.Natl.Acad.Sci.,U.S.A.76:5239-5243(1979); Gusellaら, Nature 306:234-238(1983)〕。RNAは、Taylorら, Nature Genet. 3:223-227(1992)に記載されているようにして調製し、ノーザンプロットティングを行った。

【0069】コスミド整列群の構築： コスミド整列群の最初の構築は、コスミドL19及びBJ56〔Allittoら, Genomics 9:104-112(1991); Linら, Somat.Cell Mol.Genet.17:481-488(1991)〕からの染色体ウォーキング(染色体歩行)に基づくものであった。2つのライブラリーを使用した：還元細胞ハイブリッドH39-8C10〔Whaleyら, Som.Cell Mol.Genet. 17:83-91(1991)〕由来のAlu陽性コスミドのコレクション、及びロスアラモス・ナショナル・ラボラトリーから提供された整列化フロー選別染色体4コスミドライブラリー(arrayed flow-sorted chromosome 4 cosmid library)(NM87545)。ウォーキングは、反復配列及びベクター配列のサブレーションを用いて、全コスミドDNAとロボット作成した高密度フィルター・グリッド〔Nizetc, D.ら, Proc.Natl.Acad.Sci.,U.S.A.88:3233-3237(1991); Lehrach,H.ら, in Genome Analysis: Genetic and Physical Mapping, 1巻, Davies,K.E.ら, 編, コールド・スプリング・ハーバー・ラボラトリー・プレス, 1991, 39-81〕とのハイブリダイゼーションによって行った。コスミドL1C2、L69F7、L228B6及びL83D3は、YACクローンYGA2と同じ整列ライブラリーとのハイブリダイゼーションによって最初に同



定された [Batesら, *Nature Genet.* 1:180-187(1992); Baxendaleら, *Nucleic Acids Res.* 19:6651(1991)]。HDコスミドGUS72-2130は単一コピープローブを用いるGUS72コスミドライブラリーの標準的スクリーニングによって単離された。コスミドの重複は、クローンとクローン及びクローンとゲノムハイブリダイゼーションの組合わせ、単一コピープローブハイブリダイゼーション及び制限マッピングによって確認した。

【0070】cDNAの単離及び特性化: Bucklerら, *Proc. Natl. Acad. Sci., U.S.A.* 88:4005-4009(1991)に記載されているように、エキソンプローブを単離し、クローンした。エキソンプローブ及びcDNAを使用し、成人前頭皮質、胎児脳、アデノウイルス形質転換網膜セルラインRCA及び肝RNAから構築されたヒトラムダZAP II cDNAライブラリーをスクリーニングした。cDNAクローン、PCR産物及びトラップ化エキソンはSangerら, *Proc. Natl. Acad. Sci., U.S.A.* 74:5463-5467(1977)に記載されているようにして配列決定した。直接コスミド配列決定はMcClatcheyら, *Hum. Mol. Genet.* 1:521-527(1992)に記載されているようにして行った。データベースの検索は、National Center for Biotechnology InformationのBLASTネットワークサービスを使用して行った [Altschulら, *J. Mol. Biol.* 215:403-410(1990)]。

【0071】(CAG)<sub>n</sub> 反復のPCR検定: (CAG)<sub>n</sub> 反復にフランキングするゲノムプライマー (配列番号3及び配列番号4) は次のものである:

5' ATG AAG GCC TTC GAG TCC CTC AAG TCC TTC 3'、及び

5' AAA CTC ACG GTC GGT GCA GCG GCT CCT CAG 3'

PCR増幅は、ゲノムDNA 50 ng、各プライマー5 µg、10 mM トリス-HCl pH 8.3、5 mM KCl、2 mM MgCl<sub>2</sub>、200 µM dNTPs、10% DMSO、0.1単位パーフェクトマッチ(Perfectmatch) [ストラテジーン]、2.5 µCi <sup>32</sup>P-dCTP [アメルシャン] 及び1.25単位Taqポリメラーゼ [ペーリンガー・マンハイム] を用い、反応容量25 µl 中で行った。94°Cに1.5分加熱した後、反応ミックスを以下のプログラムに従ってサイクルさせた: 40 ×

[1'@94°C; 1'@60°C; 2'@72°C]。各PCR反応物5 µl を等量の95%ホルムアミド・ローディング色素で希釈し、95°Cで2分間、熱変性する。得られた産物を5%変性ポリアクリルアミドゲル上で解析する。鋳型としてコスミドL191F1 (CAG<sub>18</sub>) を用いて行ったこの反応のPCR産物は247 bp であっ

た。対立遺伝子のサイズを、DNA配列決定はしど、はしど決定したコスミド由来のPCR産物、及びゲル上に存在することの多い不変のバックグランド・バンドと比較して算定した。対立遺伝子の変異の算定は、無関係な巨大西欧州祖先の個体及び種々の系統由来の罹患HD個

体の正常な親をタイピング (型別) することで行った。

【0072】実施例5-8におけるHD及び正常染色体の型別: HD染色体は、徴候性個体及び連鎖マーカー分析によって遺伝子のキャリアーであることが知られている「危険性ある」個体から誘導した。すべてのHD染色体は、遺伝子連鎖及び平衡異常の研究に既に用いられている種々の民族背景を有する十分に特性化されたHDファミリーの一員から得た [MacDonald, M.E.ら, *Nature Genet.* 1:99-103(1992); Conneally, P.M.ら, *Genomics* 5:304-308(1989)]。使用した150ファミリーのうち3つが巨大な系統であり、それぞれ単一の祖先からその系統を引くものであった。巨大ベネズエラHD系統は13,000員にのぼる膨大な血族であり、そこから我々は75のHD染色体を型別した [Gusella, J.F.ら, *Nature* 306:234-238(1983); Wexler, N.S.ら, *Nature* 326:194-197(1987)]。Zファミリー及びDファミリーとして従来記載されている2つの他の巨大なファミリーはそれぞれ25及び35 HD染色体を与えた [Folstein, S.E.ら, *Science* 229:776-779(1985)]。正常な染色体は、HDファミリーの既婚者(married-ins)及び非HDファミリーの無関係な正常個体から採取した。4つ以外のすべての個体について検定したDNAは、リンバ芽球セルライン又は新鮮血から調製した [Gusella, J.F.ら, *Nature* 306:234-238(1983); Anderson及びGusella *In Vitro* 20:856-858(1984)]。例外のケースでは、DNAは凍結小脳から調製した。リンバ芽球、新鮮血又は脳のDNA間におけるPCR産物の特性には相違は認められなかった。年齢24-30のベネズエラ系統の5つの員についても、本発明者らは精液試料からベレット化精子を抽出することでDNAを調製した。すべてのDNAのHD遺伝子(CAG)<sub>n</sub> 反復の長さをポリメラーゼ鎖反応増幅によって評価した。

【0073】実施例5-8に記載する数学的分析: 反復の長さとは症状発現の年齢との関係をPearson相関係数及び多変量回帰によって評価し、より高次の関係を評価した。すべてのHD染色体の反復の長さの分布と個体ファミリーのそれとの比較を、両群間の分散及びt検定分析によって行った。95%信頼バンドを、SAS [SASインスティテュートInc., SAS/STAT ユーザーガイド, バージョン6, 4編, 2巻(SASインスティテュートInc.カーリー, N.C., 846頁, 1989)] の一般線形モデル操作を利用する回帰線付近と見積もった。

【0074】実施例1

トラップクローン化エキソンを入手するためのエキソン増幅の適用

十分に特性化されたファミリーの不連続な組換え事象によって規定されるHD候補領域は図1に示されるようにD4S10及びD4S98間の2.2 Mb にまたがっている。D4S180及びD4S182間の500 kb セグメントは最も強いHDとの連鎖平衡異常を示し、共通

する単相型を共有する疾患染色体の約1/3がD4S127及びD4S95における多重対立遺伝子多形によってつなぎ止められて[MacDonaldら, Nature Genet. 1:99-103(1992)]。D4S180からD4S95及びD4S182間の位置までの約480 kb をまたぐ64重複コスミドを単離するに当たり、YAC [Baxendaleら, Nucleic Acids Res. 19:6651(1991)]からの情報及び、染色体4特異的なライブラリー及びさらにはこの領域を網羅する付加的なライブラリーの高密度フィルター・グリッドとのコスミドブローハイブリダイゼーションからの情報の組合わせ、に基づいて行った。完全な整列群を付与するこれらのコスミドの16個を図1に示す。我々は以前、エキソン増幅を使用し、D4S127とは離れた領域のADDA、 $\alpha$ -アデュシン(adducin)座、IT10C3、新規な推定トランスポーター(運搬体)遺伝子、及びIT11、新規なGプロテインカップリング化レセプターキナーゼ遺伝子を同定した(図1)。

【0075】本発明者らはここにエキソン増幅手法を、D4S127に隣接する整列群の領域由来のコスミドに適用した。この手法によりトラップ化エキソンクローンが得られ、それは単一エキソン、又は一緒にスプライスする多重エキソンの場合があり、この手法はcDNAライブラリーをスクリーニングするためのブローブを入手する効率的な方法である。個々のコスミドをプロセッシングし、コスミドL134B9からL181B10までの領域の9つのエキソンクローンを得た。エキソンブローブを用いて2つの非重複cDNAが最初に単離された。形質転換成人網膜細胞cDNAライブラリーをエキソンクローンDL118F5-Uを用いてスクリーニングし、IT15Aが入手された。IT16Aは、3つのエキソンクローン、DL83D3-8、DL83D3-1及びDL228B6-3のブールを用いて成人前頭皮質cDNAライブラリーをスクリーニングすることにより単離された。ノーザンブロット分析により、本発明者らはIT15A及びIT16Aが実際、同じ大きさの約10-11 kb 転写体の別個の部分であることを発見した。図2は、正常な個体、及び別種の単相型のHD染色体の2つの独立したホモ接合体であるリンバ芽球セルライン由来のRNAを含有するノーザンブロットの例を示すものである。同じ約10-11 kb 転写体が、種々のヒト組織(肝、脾臓、腎臓、筋肉、及び成人脳由来の種々の領域)由来のRNAにも検出された。

【0076】IT15A及びIT16Aを使用し、完全長の転写体を入手するため、多くのヒト組織cDNAライブラリーにてウォーキングさせた。図3は、IT15転写体を規定する5つのcDNAクローンを、凡例で説明するように誘導した混成配列の模式図の下に示している。IT15の混成配列は全ペプチドコード化配列を含有しており、図4、図5及び図6に示しているように18Aの尾を包含する10,366塩基をまたいでいる。

9,432塩基のオープン解読枠は316塩基の可能性あるイニシエーターメチオニンコドンから始まり、それは最適な翻訳開始配列に関連して配置されている。枠内終止コドンはこの部位の240塩基上流に位置している。IT15のタンパク質産物は、3,144アミノ酸を含有する348 kDaタンパク質と推定される。長いオープン解読枠内の最初のMetコドンは可能性あるイニシエーターコドンとして選択したが、我々は、翻訳が実際にはもっと3'側のMetコドンから始まっておらず、より小さなタンパク質が産生されるという事実は排除しきれなかった。

#### 【0077】実施例2

##### (CAG)<sub>n</sub> トリヌクレオチド反復の多形変異

その5'末端に近い、IT15配列はグルタミンをコードしているCAGトリプレットの21コピーを含有している(図5)。この配列を、4p16.3における単純配列反復(SSR)の囲むことが知られているゲノム配列と比較すると、正常なコスミドL191F1は(CAG)<sub>n</sub> 反復が多形であることを示す18コピーのトリプレットを有していることが判明した(図5)。この反復にフランキングするゲノム配列由来のプライマーを選択し、この変異のためのPCR検定を確立した。正常な集団では、このSSR多形は、約11から約34の反復単位である少なくとも17個の別個の対立遺伝子(表1)を示した。173の正常染色体の98%が11-24反復の反復長さを有していた。2つの染色体は25-30の反復範囲で検出され、2つの正常染色体はそれぞれ33及び34の反復を有していた。正常な染色体上の異種接合体の全体は80%であった。3つのクローンの配列決定分析によれば、この変異は完全に(CAG)<sub>n</sub>に基づくものであると考えられるが、比較的小さな下流の(CCG)<sub>n</sub>の変異(これはPCR産物に包含されているが)の可能性もある。

#### 【0078】実施例3

##### HD染色体上のトリヌクレオチド反復の不安定性

主要なHD単相型を有する染色体由来のコスミドGUS72-2130の配列決定分析により、最も大きな正常対立遺伝子よりも遥かに大きなトリヌクレオチド反復の48コピーが判明した(図7)。PCR検定をHD染色体に適用する場合、正常の変異とは著しく相違するパターンが認められた。HD異種接合体は正常なサイズ範囲の1つの独立した対立遺伝子産物、及びそれよりも大きなサイズの1つのPCR産物を含有しており、そのことはHD染色体上の(CAG)<sub>n</sub> 反復が正常な染色体と相関して伸長していることを示している。

【0079】図8は、PCR検定を巨大ベネズエラHD血縁の選択した核家族由来のリンバ芽球DNA上で行ったときに観察されるパターンを示している。この家族では、従前のDNAマーカー分析はHD染色体が父親(レーン2)から7人の子供(レーン3、5、6、7、8、



10及び11)に伝達されたことを示していた。この交配に存在する3つの正常染色体は、メンデルスの法則の態様で遺伝する正常サイズにあるPCR産物を与えた

(AN1、AN2、AN3)。父親のHD染色体は、非ベネズエラHDコスミドの48反復産物よりも若干小さな拡散した「ファジー(ぼやけ)」様のPCR産物を与えた。PCR増幅しなかったレーン5のDNA及び単一の正常対立遺伝子のみを示すレーン11のDNAを除き、罹患子供のDNAそれぞれは別個のサイズのファジーなPCR産物(AE)を与え、このことはHD染色体(CAG)<sub>n</sub>反復が不安定であることを示している。レーン6は、父親のDNAの産物よりも若干小さな、又は同等のHD特異的な産物を含有している。レーン3、7、10及び8はそれぞれ、徐々にサイズが大きくなったHD特異的PCR産物を含有していた。レーン11にはHD特異的なPCR産物が存在しないが、それはこの子供のDNAが、長過ぎて効率的に増幅されない(CAG)<sub>n</sub>反復を有していることを示している。これはサザンブロット分析によって確認され、ここでは伸長HD対立遺伝子が容易に検出され、100までの反復コピーを含有すると評価された。特に、この子供は非常に早期な年令の2才でHDを若年期発症した。父親のHDの発症は彼にとっては早期の40才であり、この集団の殆どの成人HD患者に普通の年令であった。レーン3、7、10及び8で示されるこの子供の発症年令はそれぞれ26、25、14及び11才であり、このことはHDの発症年令とHD染色体上の(CAG)<sub>n</sub>反復の長さとの間に大まかな相関性があることを示している。最も少ない反復を有するレーン6で示される子供は、この傾向と異なることなく、年令30才で最後に検査しても前徴候を保持していた。

【0080】図9は、両親がDNAマーカー試験に基づいて同じHD染色体を担持するHD異種接合体であるベネズエラ系統由来の第2の同胞群のPCR分析を示している。幾つかの子供は既に報告されているようにHDホモ接合体(レーン6+7、10+11、13+14、17+18、23+24)である[Wexlerら, Nature 326:194-197(1987)]。各親のDNAはメンデル法則の態様で伝達される正常範囲(AN1、AN2)の1つの対立遺伝子を含有していた。両親及び子供のDNA由来のHD特異的な産物(AE)はすべて、正常な対立遺伝子の産物よりも十分に長く、また平均サイズに膨大な変動を示した。現在進行中のベネズエラHDプロジェクトに関連する研究者の盲検状態を維持するため、この系統の子孫の神経学的診断は行わなかったが、発症の年令はここでも反復の長さと平行しているようである。多くの個々の印の対合サンプルは、少なくとも1年後に開始した独立のリンバ芽球ラインを表している。対合サンプル間の変動は異なる個体間程には大きくなく、このことはPCR産物のサイズの主要な相違が減数分裂の伝達に由来

するものであることを示している。特に留意すべきは、レーン13及び14で得られた結果である。このHDホモ接合体のDNAは両親のHD特異的なPCR産物よりも大きな1つのPCR産物と小さな産物とを与えた。

【0081】現在までのところ、我々は、MacDonaldら, Nature Genet. 1:99-103(1992)にてすべて別個に、そして広範な民族背景で報告されている75の独立したHDファミリーを試験している。75の症例すべてが、正常サイズの範囲よりも大きなPCR産物をHD染色体から産生した。HD特異的な産物のサイズは、42反復コピーから66コピーよりも大きなサイズの範囲であるが、2、3の個体は極めて長い反復のために産物を与えなかった。これらの症例では、サザンブロット分析により、100コピーの反復に近い最も巨大な対立遺伝子を有するEcoRI断片の長さが増大していることが判明した。図10は、主要なHD単相型が分離されているアイルランド人を祖先とするアメリカ・ファミリーの一員に検出された変動を表している。レーン2のDNAを増幅したHDホモ接合体から、コスミドGUS72-2130をクローンした。異なる4p16.3単相型によって障害が分離しているベネズエラHD系統にて観察されているように(図6及び図7)、このファミリーのHD特異的なPCR産物は相当なサイズ変動を示している。

#### 【0082】実施例4

##### HDに対する新規な突然変異

HDでの突然変異率は非常に遅いと報告されている。(CAG)<sub>n</sub>反復の拡大が新たなHD突然変異の出現による機構であるのか否かを試験するため、膨大な調査によって障害の家族歴を究明できなかった、HDの散発症状を有する2つの系統を調査した。これらの症例では、罹患個体と非罹患の親類との両方で同じ染色体を同定するのに十分な系統情報を集めた。図11及び図12には、これらのファミリーの(CAG)<sub>n</sub>反復のPCR分析の結果を示している。別個の単相型を規定する4p16.3の膨大な数のRFLP及びSSRマーカーの型別に基づき、各ファミリーの染色体に任意の数字を帰属し、推定HD染色体に星印を付ける。

【0083】ファミリー#1では、HDは、障害が染色体3\*と共にIII-1に伝達した個体II-3に最初に現れた。この同じ染色体は、高齢で罹患していない個体であるII-2に存在した。PCR分析により、II-2由来の染色体3\*は正常範囲の極めて高い末端でPCR産物を産生することが判明した(約36CAGコピー)。しかし、II-3及びIII-1における同じ(CAG)<sub>n</sub>反復はそれぞれ約44及び約46コピーまで連続して拡大した。同様の結果がファミリー#2で得られ、ここでは、推定のHD突然変異体III-2がII-1及びIII-1の同じ染色体に比較して相当に拡大した反復を有していた(約49に対し、約33CAGコピー)。ファミリー#1及びファミリー#2共に、最終的HD染色体がHD染

色体すべての1/3のマーカー単相型特性を示し、そのことはこの単相型が反復の拡大に陥り易くなっているかもしれないことを表している。

#### 【0084】議論

IT15遺伝子内のHD染色体上における拡大し不安定であるトリヌクレオチド反復の発見は、この遺伝子を本発明のHD遺伝子として利用できることを基礎としている。これらの結果は、HDは、ヒト遺伝子疾患に全く共通していることが証明できる突然変異メカニズムの最新の例であるという解釈と矛盾がない。トリヌクレオチド反復配列の伸長は3つの全く異なるヒト疾患、脆X症候群、筋緊張性ジストロフィー及び脊髄延髄性筋萎縮の原因として以前から関与している。反復の拡大がHDにて最初に観察されたことから、この現象はこれらの各障害と共通の特徴を共有していることが判明した。

【0085】脆X症候群では、精神薄弱などの徴候の配列及びXq27.3での脆弱部位の発現は、FMR1遺伝子の5'非翻訳領域にあると考えられる(CGG)<sub>n</sub>反復の拡大に関連している[Fuら, Cell 67:1047-1058(1991); Kremerら, Science 252:1711-1714(1991); Verkerkら, Cell 65:904-914(1991)]。筋緊張症を伴う筋肉脆弱化が関与する通常は成人初期に現れる優勢な障害である筋緊張性ジストロフィーでは、不安定なトリヌクレオチド反復:(CTG)<sub>n</sub>がミソトニン(myotonic)プロテインキナーゼ遺伝子の3'非翻訳領域内に位置している[Aslanidisら, Nature 355:548-551(1992); Brookら, Cell 68:799-808(1992); Buxtonら, Nature 355:547-548(1992); Fuら, Science 255:1256-1259(1992); Harleyら, Lancet 339:1125-1128(1992); Mahadevanら, Science 255:1253-1255(1992)]。HDにおける不安定な(CAG)<sub>n</sub>反復はIT15遺伝子のコード化配列内にあることができ、男性ホルモンレセプター遺伝子のコード化配列内の(CAG)<sub>n</sub>反復の拡大が原因である運動神経のX連結劣性成人発症障害である脊髄延髄性筋緊張症と、その特徴が共通している[LaSpadaら, Nature 352:77-79(1991)]。脆X症候群及び筋緊張性ジストロフィーの両者における反復の長さは、継承世代を増加する傾向にあり、それは時には極めて劇的である。平均の反復長さの減少が時折観察される場合がある[Fuら, Science 255:1256-1259(1992); Yuら, Am. J. Hum. Genet. 50:968-980(1992); Grunerら, N. Engl. J. Med. :476-480(1993)]。HDトリヌクレオチド反復も不安定であり、次世代に伝達されるときに拡大するのが通常であるが、時には縮まる場合もある。HDでは、他の障害と同様に、組換えが存在しなくてもコピー数の変動が起こる。脆X症候群、筋緊張性ジストロフィー、及びHDを比較すると、脊髄延髄性筋緊張症の疾患対立遺伝子の不安定性はより制限されており、反復の長さの劇的な拡大は認められない[Biancalanaら, Hum. Mol. Genet. 1:255-258(1992)]。

【0086】筋緊張性ジストロフィーの反復の長さの拡大

大は特定の染色体単相型と関連しており、これは原始の容易突然変異(primordial predisposing mutation)の存在を示唆している[Harleyら, Am. J. Hum. Genet. 49:68-75; Harleyら, Nature 355:545-546(1992); Ashizawa, Lancet 338:642-643(1991); 及びEpstein (1991)]。脆X症候群では、トリヌクレオチド反復の数が増大し易い先祖伝来の突然変異の限定された数が存在し得る[Richardsら, Nature Genet. 1:257-260(1992); Oudetら, Am. J. Hum. Genet. 52:297-304(1993)]。IT15を同定するために使用される連鎖平衡異常により、HDに関連する幾つかの単相型があることが示されたが、HD染色体の少なくとも1/3は祖先に関連するものである[MacDonaldら, Nature Genet. 1:99-103(1992)]。これらのデータを、HDに対する低率の突然変異の報告[Harper, J. Med. Genet. 89:365-376(1992)]と組み合わせると、トリヌクレオチド反復の拡大は特別の染色体上で起こるのみであり得ることが示された。新たな突然変異が起こっていることが疑われる本明細書にて示している2つのファミリーの分析は、HD域での反復の拡大を受けることのできる特別な正常染色体が存在し得るという見識と矛盾が無い。これらのファミリーではそれぞれ、正常範囲の上限にある(CAG)<sub>n</sub>反復の長さを有する染色体が、これら2つの症例においてHD染色体及びHDの臨床所見に認められる最も共通する単相型に適合する4p16.3単相型がトリヌクレオチド反復の拡大と関連している染色体上で分離していた。

【0087】HD染色体上の連鎖平衡異常を調査するために単相型分析が最近適用されたが、それによりHD系統で説明される大多数の組換え事象で規定できる2.2Mb候補領域の部分が明らかになった[MacDonaldら, Nature Genet 1:99-103(1992)]。従前ならば、遺伝子の研究は、遺伝子継承のパターンが残りのファミリーと一致していない3つの交配のために複雑であった[MacDonaldら, Neuron 3:183-190(1989b); Prichardら, Am. J. Hum. Genet. 50:1218-1230(1992)]。これらの交配は、系統の残りに存在するHD染色体の継承が効率的に排除され、4p16.3全体のマーカーについて正常な対立遺伝子のみが継承されるとしても、明らかにHD罹患の個体を産した。上記のPCR検定を使用してこれらの各ファミリーを試験することで、他のHD血統のように、拡大された対立遺伝子が3つすべての系統の罹患個体でHDと分離していることを決定した。しかし、拡大された対立遺伝子は不統一な4p16.3遺伝子型を有する特定の個体には存在しなかった。これらの個体は代わりに、4p16.3内の他のマーカーの分析に基づいて正常であると期待される対立遺伝子を現した。これらの不統一な個体は実際、HDを有していないが、何らかの別の障害を有していると考えられる。あるいは、これらの個体は、HD対立遺伝子が遺伝子型別方法に使用されるリンバ芽球DNAにおけるよりも脳組織において重く現



れ、及び／又はより拡大している遺伝子モザイクを示すかもしれない。

【0088】HDについての「危険性」を有する個体においてトリヌクレオチド反復のサイズを直接モニターできることは現在の方法よりも顕著な有益性を提供するものであり、つまり複雑な連鎖分析を行う必要がなく、遺伝学カウンセリングが容易であり、生存している罹患親類を持っていない「危険な」個体に対する前徴候及び出生前診断の適用性が拡大している。しかし、危険性を有する者を試験するための現在国際的に許容されているガイドライン及びカウンセリングのプロトコールは監視され続けており、また非罹患の親類から得たサンプルは不注意で、即ち十分な同意無くして行うべきでないことは、最も重要なことである。この研究で調査されている一連の患者には、反復の長さとの間に明らかな相関性が認められ、これは筋緊張性ジストロフィーを思い起こさせる[Harlyら, Lancet 339:1125-1128(1992); Tsilfidisら, Nature Genet. 1:192-195(1992)]。最も巨大なHDトリヌクレオチド反復のセグメントは若年期発症の症例で見いだされており、ここでは雄性伝達が優位であることが知られている[Merrittら, Excerpta Medica, アムステルダム, 645-650頁(1969)]。

【0089】脆X症候群の発現にはFMR1遺伝子の直接的な不活化が関連している[Pierrettiら, Cell 66:817-822(1991); DeBouilleら, Nature Genet. 3:31-51(1993)]。脊髄延髄性筋緊張症の劣性遺伝パターンは、この障害に不活性遺伝子産物が産生されることを示している。筋緊張性ジストロフィーでは、反復の拡大が優勢な疾患表現型を導く様子は知られていない。HDのトリヌクレオチド反復の拡大の発生学的メカニズムには多くの可能性が存在する。この理論にとらわれるつもりはなく反対に、4p16.3に関して単価遺伝子接合性のウォルフ-ハーシュホルン患者はHDの特徴を現さず、IT15 mRNAはHDホモ接合体に存在することから、トリヌクレオチド反復の拡大はそれを含有する遺伝子の単純な不活化を引き起こすものではないという点に注目することができる。この疾患対立遺伝子のホモ接合体は異種接合体とは臨床的に異なっていないため、HDの表現型が完全に優性であるという観察事項により、この疾患の対立遺伝子のmRNA産物又はタンパク質産物のいずれかが何らかの新たな性質を有しており、又はそれらが不適当に発現される機能突然変異の獲得がHDの原因であると示された[Wexlerら, Nature 326:194-197(1988)] \*

\* 7); Myersら, Am. J. Hum. Genet. 45:615-618(1989)]。拡大されたトリヌクレオチド反復が翻訳されれば、タンパク質産物はN末端付近のポリグルタミン長の長さが増大し、その結果は劇的となる。しかし、上流にMetコドンが存在するにもかかわらず、正常な翻訳開始が(CAG)<sub>n</sub>反復の3'側で起こり、得られるタンパク質産物にポリグルタミン長が存在しない可能性がある。この場合、反復は5'非翻訳領域内にあり、mRNAレベルで優勢な効果を示すと予想される。反復の拡大が存在すれば、それを含有するmRNAの調節、位置決め、安定性又は翻訳性が直接的に変動し、間接的にHD異種接合体の正常な対立遺伝子由来のその対応物に影響することができる。他の考えられる科学的モデルは、反復の拡大の存在により、HD転写体の効果的な翻訳開始部位が変動することによってタンパク質を断頭し、又はIT15遺伝子の転写開始部位が変動することによってmRNAの発現制御が破壊されるというものである。最後になるが、この反復はIT15転写体内に位置しているが、それが隣接遺伝子の発現に対する作用によってHDを導くという可能性は排除できない。

【0090】この最後の考え方にかかわらず、それは上記の結果と矛盾がなく、ハンチンチンと命名したIT15遺伝子のタンパク質産物の発現及び／又は構造に対するmRNA又はタンパク質レベルでの作用に基づいてトリヌクレオチド反復の拡張がHDの原因となることは尤もらしい。トリプレット反復の領域の外側ではIT15 DNA配列は、GenBankデータベースに既に報告されている遺伝子とは有意な類似性が見いだされなかった。N末端付近のグルタミン及びプロリンの伸長を除き、そのアミノ酸配列は既知のタンパク質と類似性を示さず、ハンチンチン機能についての目だった糸口も何ら提示されていない。N末端付近のポリグルタミン及びポリプロリン領域は、これらのアミノ酸の長い伸長部を含有する大多数のタンパク質と類似性を示している。このような類似性の有意さを評価するのは困難であるが、これらの多くのものがDNA結合タンパク質であり、ハンチンチンが1,443残基にて単一のロイシン・ジッパー文様を有していることは注目に値する。ハンチンチンは広範に発現されているようであり、またHDの細胞死は脳の特定領域の特定ニューロンに限定されている。

【0091】

【表1】

HD及び正常反復サイズの比較

対立遺伝子の範囲 (#反復)	正常染色体の数及び頻度		HD染色体の数及び頻度	
≥ 48	0	0	44	0.59
42-47	0	0	30	0.41
30-41	2	0.01	0	0
25-30	2	0.01	0	0

33		34		
≤ 24	169	0.98	0	0
合 計	173	1.00	74	1.0

## 【0092】実施例5

## 正常及びHD染色体上におけるトリヌクレオチド反復の長さの分布

HDトリプレット反復のコピー数を150の独立ファミリー由来の425 HD染色体の総数で調査し、545の正常染色体のHDトリプレット反復のコピー数と比較した。得られた結果を図13に示している。反復の長さの2つの非重複分布が認められ、正常範囲の上限とHD範囲の下限は3反復単位で分かれていた。正常染色体は11から34の反復単位のPCR産物を産する24個の対立遺伝子を示し、中央値は19単位であった（平均19.71, 標準偏差3.21）。HD染色体は37から86単位の反復長に相当する54個の分離したPCR産物を与え、その中央値は45単位であった（平均46.42, 標準偏差6.68）。

【0093】HD染色体のうち134及び161はそれぞれ母親又は父親由来であることが知られている。伝達する親の性が反復長の分布に影響するか否かを調査するため、これらの2つの染色体の組みを別々に図14及び図15にプロットしている。母親由来の染色体は37から73単位の反復の長さを示し、その中央値は44であった（平均44.93, 標準偏差5.14）。父親由来の染色体は反復単位を37から86コピー有しており、その中央値は48単位であった（平均49.14, 標準偏差8.27）。しかし、父親由来のHD染色体の比較的高い比率は55単位よりも大きな反復の長さを有しており（16%対2%）、このことは父親と母親の伝達の効果が異なっている可能性を示唆している。

【0094】使用したデータの組みは、HD染色体が遺伝されていないことがDNAマーカー連鎖の研究によって既に示されている2、3の臨床的に診断した個体から得た染色体を排除している [MacDonald, M.E.ら, Neuron 3:183-190(1989); Pritchard, C.ら, Am. J. Hum. Genet. 50:1218-1230(1992)]。これらの個体は十分に正常範囲内の反復の長さを有している。彼らの疾患の発症は説明されていないが、彼らはHDの表現型形質を現すことができる。関与するメカニズムとは無関係に、診断結果が反復の長さのみに基づく場合は既知のHDファミリー内のこのような個体の発現頻度は低いと考えねばならない。

【0095】対照データの組みも、HDの「自発的」症例又は「新たな突然変異」に関連する表現型的に正常な個体から得られる多くの染色体を排除している。臨床的には罹患されておらず、かつ障害の家族歴を持たないこれらの個体由来の染色体はHDと帰属することはできない。しかし、これらの染色体は「自発的」HD発端者と帰属される罹患した親類のものと反復配列を除いて実質的に同じ染色体であるので、その染色体は明確に正常と

も分類できない。これらの曖昧な染色体上に見いだされる反復の長さ（34-38単位）は対照及びHD分布間の間隙をつなぐものであり、低いHD範囲に対して高い正常な反復を有している個体の状態は決定するのが困難となっている。

## 【0096】実施例6

## トリヌクレオチド反復の不安定性

図13に示すデータは、欠損の多くの可能性ある独立起源を示す別個の150のHDファミリー由来の反復の長さをまとめたものである。共通の子孫から継承されていることが知られているHD染色体の組みの反復の長さにある変異を調べるため、別個の4p16.3単相型を有する3つの巨大HD血統 [Gusella, J.F.ら, Nature 306:234-238(1983); Wexler, N.S.ら, Nature 326:194-197(1987); Folstein, S.E.ら, Science 229:776-779(1985)] であってそれぞれ75、25及び35個体を型別（タイピング）したものを分離した。各系統内の子孫HD染色体が単一の起源であるにもかかわらず、別の系統の一員は広範な反復の長さを示している [図16、図17、図18]。HD染色体の反復の不安定性は巨大ベネズエラHD血統の一員において最も優性である（図16）。[共通するHDの祖先は10世代の子孫を産し、数にして13,000個体に及ぶ]。ベネズエラ系統のこのサンプリングにおける反復長さの分布（中央値46, 平均48.26, 標準偏差9.3）は、すべてのファミリー由来のHD染色体のもっと大きなサンプルから得られたものと有意に異なっていない。図17及び図18は、HDがベネズエラ血統におけるよりも最近になって導入されている2つの拡大ファミリーの結果を示している。これらのファミリーは異なる発症年令の分布を示し、HDの表現型の特徴が変動していると報告されている [Folstein, S.E.ら, Science 229:776-779(1985)]。この両者により、反復長さの大きな変動が明らかになり、それぞれ中央値が40及び49反復単位であった。図17のファミリーの一員における反復長さの分布はすべてのHD染色体の反復長さの分布と有意に相違しており（ $p < 0.0001$ ）、小さいほうの平均は42.04反復単位であった（標準偏差2.82）。図18のHD染色体由来の反復の分布も全データの組み由来のものと有意に相違していたが（ $p < 0.004$ ）、高いほうの平均49.80（標準偏差5.86）であった。

## 【0097】実施例7

## 反復の長さ変動に対する父親供給源の効果

図13の62 HD染色体では、トリヌクレオチド反復の長さも、相当する親のHD染色体上で調査することができた。25の母親伝達のうちの20、及び37の父親伝達のうちの31において反復長さが改変しており、これは相当な不安定性を示している。500以上の減数分裂



の伝達により反復長さに変化が認められない正常な染色体では、類似の現象は観察されなかったが、このような大多数の正常な対立遺伝子が本当に存在することは少なくとも低い程度の不安定性を示すものである。

【0098】図19及び図20は、罹患した親と対応する子供のHD染色体上の反復長さ間の相関性を示している。反復の長さが改変している20の母親遺伝子的染色体では、13の変化が長さの増大であり、7つが減少であった。増大及び減少ともに5つの反復単離よりも少ない変化が関与しており、母親の反復の長さとその母親の子供のそれとの相関性は $r = 0.95$  ( $p < 0.0001$ )であった。25の母親伝達における反復長さの平均の変化は0.4反復の増大であった。

【0099】父親由来の染色体では、反復長さが変化している31の伝達の内訳は26の長さ増大と5の長さ減少であった。このサイズの減少は母親由来の染色体で観察されるよりも若干しか、即ち1-3反復単位の範囲でしか小さくないが、その増大は劇的に大きいことが多い。従って、父親の反復長さと彼の子供のそれとの相関性は $r = 0.35$  ( $p < 0.04$ )しかなかった。37の父親伝達における平均の変化は9反復単位の増大であった。父親伝達で観察された最大の長さ増大は41反復単位であり、これは父親反復の2倍に近かった。雄性及び雌性伝達ともに、父親反復のサイズと変化の強さ及び頻度いずれかとの間には相関性は無かった。

【0100】HD染色体の雄性伝達で観察される反復の長さの変動が雄性の胚細胞に反映されるか否かを決定するため、我々は5HD遺伝子キャリアーから得た精子DNA及び対応するリンパ芽球のDNA由来の反復を増幅した。図21に示す結果は、HD染色体の反復についてはリンパ芽球及び精子DNA間に顕著な相違があるが、正常染色体の反復には存在しないことを示している。精子ドナーのすべてはベネズエラHDファミリーの一員であり、24から310才のヒトであった。個体1及び2はリンパ芽球DNAに基づくHD染色体の反復長さをそれぞれ45及び52で有している同胞である。個体3及び4はHD反復長さがそれぞれ46及び49である、これらも同胞である。個体5は別の2対のいずれかでない別の同胞由来であるが、52コピーのHD反復を有している。5症例はすべて、精子DNA及びリンパ芽球DNAのPCR増幅により、正常染色体由来の産物と同じものを与えた。しかし、リンパ芽球DNAと比較すると、精子DNA由来のHD遺伝子は拡散配列の産物を与えた。5症例のうち3つ(2、4および5)は、拡散配列が対応するリンパ芽球産物よりも極めて大きな対立遺伝子産物にまで広がっている。被験者2は最も大きな程度の拡大を示し、その精子DNA産物は80反復単位にまで及んでいる。興味深いことに、最も大きな変動を示す個体3は最も長い反復を有しており、現在徴候性であった。別の2人のドナーはそれよりも短い長さのHD範囲

の反復を有し、この時点では危険なままである。

【0101】父親から伝達されたHD染色体と母親から伝達されたそれとの間の反復長さが高い範囲で顕著に相違していることは、親供給源効果の可能性を示している。これを直接解析すると、HD染色体の反復の長さは約85%の伝達が変動していた。殆どの変動は数個の反復単位のみ機能が関与しており、雄性伝達にしか大きな増大は起こらない。この雄性伝達における大きなサイズの増大は、精子DNAからの反復の増大に基づく、雄性の配偶子発生の際のHDトリヌクレオチド反復の特別な不安定性によって引き起こされるらしい。

#### 【0102】実施例8

##### 反復の長さ及び発症年齢間の関係

反復の長さが増大することはHDの発症年齢の低下と相関しているかもしれない。従って、発症年齢のデータを図13に示す234人の個体について測定した。図22は、HD及び正常染色体に見いだされる、発症年齢に対する反復の長さを表している。実際、発症年齢はHD反復の長さに反比例している。発症年齢と反復の長さとは一次関数の関係にあると仮定して、Pearsonの相関係数として $r = -.75$ ,  $p < 0.0001$ が得られた。多項式関数を用いた場合により良好な一致が得られたが( $R^2 = 0.61$ ,  $F = 121.45$ )、このことは発症年齢と反復長さとの間にはより高次元の関係にあることを示唆するものである。

【0103】反復単位の数と関連する発症年齢には相当な変動があり、発症が15-75才である場合の37-52単位域のトリヌクレオチド反復では特にそうである(HD染色体88%)。この範囲では、発症年齢と反復の長さとの間の一次関数に、より高次元の相関と同じ適合性が得られた。予想回帰線の回りの95%信頼区間が $\pm 18$ 才で算定された。52単位の範囲のうち37では、発症年齢に対する反復長さの関係は全分布の半分の強さしかなく( $r = -0.40$ ,  $p < 0.0001$ )、このことは52単位以上の反復が多く予想検出力に寄与していることを示している。この増大した範囲では、発症年齢は非常に若いようであり、その結果、試験を求める殆どのヒトにとっては相当でない。

【0104】HD遺伝子の親起源に基づくデータの組みを細分することのできる37-52反復単位の範囲にある178症例では、多変量回帰分析により、この範囲の反復の長さと独立した発症年齢に対する親起源の有意な効果が示された( $p < 0.05$ )。母親伝達由来のHD遺伝子キャリアーは、父親伝達由来のものより平均発症年齢が2才遅かった。単変量及び多変量分析の両者共に、全データ組み、又は母親又は父親起源の染色体に細分化した場合のいずれにおいても正常な染色体上の反復の長さとは発症年齢との間には何らの関係も検出されなかった。

【0105】上述の全文献は引用によって本明細書に包

含まれる。本発明を明瞭にし理解するため幾つかの詳細を先に記載したが、当業者ならば、本発明及び添付の請求の範囲の真の範囲を逸脱することなく、形態及び詳細の種々の変更が可能であることは理解されよう。

【0106】

【配列表】

CGCGGGAGAC CGCCATGGCG

【0108】配列番号：2

配列の長さ：17

配列の型：核酸

AATACGACTC ACTATAG

【0109】配列番号：3

配列の長さ：30

配列の型：核酸

ATGAAGGCCT TCGAGTCCCT CAAGTCCTTC

【0110】配列番号：4

配列の長さ：30

配列の型：核酸

AAACTCACGG TCGGTGCAGC GGCTCCTCAG

【0111】配列番号：5

配列の長さ：10366

配列の型：核酸

鎖の数：一本鎖

\*

\*【0107】配列番号：1

配列の長さ：20

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

※10

※鎖の数：一本鎖

トポロジー：直鎖状

★

★鎖の数：一本鎖

トポロジー：直鎖状

☆

☆鎖の数：一本鎖

トポロジー：直鎖状

20◆トポロジー：直鎖状

配列の特徴：

(A) 特徴を表す記号：CDS

◆

(B) 存在位置：316..9748

TTGCTGTGTG AGGCAGAACC TGCGGGGGCA GGGCGGGCT GGTTCCTCGG CCAGCCATTG	60
GCAGAGTCCG CAGGCTAGGG CTGTCAATCA TGCTGGCCGG CGTGGCCCCG CCTCCGCCGG	120
CGCGGGCCCCG CCTCCGCCGG CGCAGTCTG GGACGCAAGG CGCCGTGGGG GCTGCCGGGA	180
CGGGTCCAAG ATGGACGGCC GCTCAGGTTT TGCTTTTACC TGCGGCCAG AGCCCCATTC	240
ATTGCCCCGG TGCTGAGCGG CGCCCGGAGT CGCCCCGAGG CCTCCGGGGA CTGCCGTGCC	300
CGGCCCGAGA CCGCC ATG GCG ACC CTG GAA AAG CTG ATG AAG GCC TTC GAG	351
Met Ala Thr Leu Glu Lys Leu Met Lys Ala Phe Glu	
1 5 10	
TCC CTC AAG TCC TTC CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	399
Ser Leu Lys Ser Phe Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln	
15 20 25	
CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAA CAG CCG CCA CCG CCG	447
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Pro Pro Pro Pro	
30 35 40	
CCG CCG CCG CCG CCG CCT CCT CAG CTT CCT CAG CCG CCG CCG CAG GCA	495
Pro Pro Pro Pro Pro Pro Pro Gln Leu Pro Gln Pro Pro Pro Gln Ala	
45 50 55 60	
CAG CCG CTG CTG CCT CAG CCG CAG CCG CCC CCG CCG CCC CCG CCG	543
Gln Pro Leu Leu Pro Gln Pro Gln Pro Pro Pro Pro Pro Pro Pro Pro	
65 70 75	
CCA CCC GGC CCG GCT GTG GCT GAG GAG CCG CTG CAC CGA CCA AAG AAA	591
Pro Pro Gly Pro Ala Val Ala Glu Glu Pro Leu His Arg Pro Lys Lys	
80 85 90	
GAA CTT TCA GCT ACC AAG AAA GAC CGT GTG AAT CAT TGT CTG ACA ATA	639
Glu Leu Ser Ala Thr Lys Lys Asp Arg Val Asn His Cys Leu Thr Ile	
95 100 105	
TGT GAA AAC ATA GTG GCA CAG TCT GTC AGA AAT TCT CCA GAA TTT CAG	687



39	40
Cys Glu Asn Ile Val Ala Gln Ser Val Arg Asn Ser Pro Glu Phe Gln	
110 115 120	
AAA CTT CTG GGC ATC CCT ATG GAA CTT TTT CTG CTG TGC AGT GAT GAC	735
Lys Leu Leu Gly Ile Ala Met Glu Leu Phe Leu Leu Cys Ser Asp Asp	
125 130 135 140	
CCA GAG TCA GAT GTC AGG ATG GTG GCT GAC GAA TGC CTC AAC AAA GTT	783
Ala Glu Ser Asp Val Arg Met Val Ala Asp Glu Cys Leu Asn Lys Val	
145 150 155	
ATC AAA GCT TTG ATG GAT TCT AAT CTT CCA AGG TTA CAG CTC GAG CTC	831
Ile Lys Ala Leu Met Asp Ser Asn Leu Pro Arg Leu Gln Leu Glu Leu	
160 165 170	
TAT AAG GAA ATT AAA AAG AAT GGT GCC CCT CCG AGT TTG CGT GCT CCC	879
Tyr Lys Glu Ile Lys Lys Asn Gly Ala Pro Arg Ser Leu Arg Ala Ala	
175 180 185	
CTG TCG AGG TTT GCT GAG CTG GCT CAC CTG GTT CCG CCT CAG AAA TGC	927
Leu Trp Arg Phe Ala Glu Leu Ala His Leu Val Arg Pro Gln Lys Cys	
190 195 200	
AGG CCT TAC CTG GTG AAC CTT CTG CCG TGC CTG ACT CGA ACA AGC AAG	975
Arg Pro Tyr Leu Val Asn Leu Leu Pro Cys Leu Thr Arg Thr Ser Lys	
205 210 215 220	
AGA CCC GAA GAA TCA GTC CAG GAG ACC TTG GCT GCA GCT GTT CCC AAA	1023
Arg Pro Glu Glu Ser Val Gln Glu Thr Leu Ala Ala Ala Val Pro Lys	
225 230 235	
ATT ATG GCT TCT TTT GGC AAT TTT GCA AAT GAC AAT GAA ATT AAG GTT	1071
Ile Met Ala Ser Phe Gly Asn Phe Ala Asn Asp Asn Glu Ile Lys Val	
240 245 250	
TTG TTA AAG GCC TTC ATA GCG AAC CTG AAG TCA AGC TCC CCC ACC ATT	1119
Leu Leu Lys Ala Phe Ile Ala Asn Leu Lys Ser Ser Ser Pro Thr Ile	
255 260 265	
CGG CCG ACA GCG GCT CGA TCA GCA GTG AGC ATC TGC CAG CAC TCA AGA	1167
Arg Arg Thr Ala Ala Gly Ser Ala Val Ser Ile Cys Gln His Ser Arg	
270 275 280	
AGG ACA CAA TAT TTC TAT AGT TGG CTA CTA AAT GTG CTC TTA GGC TTA	1215
Arg Thr Gln Tyr Phe Tyr Ser Trp Leu Leu Asn Val Leu Leu Gly Leu	
285 290 295 300	
CTC GTT CCT GTC GAG GAT GAA CAC TCC ACT CTG CTG ATT CTT GGC GTG	1263
Leu Val Pro Val Glu Asp Glu His Ser Thr Leu Leu Ile Leu Gly Val	
305 310 315	
CTG CTC ACC CTG AGG TAT TTG GTG CCC TTG CTG CAG CAG CAG GTC AAG	1311
Leu Leu Thr Leu Arg Tyr Leu Val Pro Leu Leu Gln Gln Gln Val Lys	
320 325 330	
GAC ACA AGC CTG AAA GGC ACC TTC GGA GTG ACA AGG AAA GAA ATG GAA	1359
Asp Thr Ser Leu Lys Gly Ser Phe Gly Val Thr Arg Lys Glu Met Glu	
335 340 345	
GTC TCT CCT TCT GCA GAG CAG CTT GTC CAG GTT TAT GAA CTG ACG TTA	1407
Val Ser Pro Ser Ala Glu Gln Leu Val Gln Val Tyr Glu Leu Thr Leu	
350 355 360	
CAT CAT ACA CAG CAC CAA GAC CAC AAT GTT GTG ACC CGA GCC CTG GAG	1455
His His Thr Gln His Gln Asp His Asn Val Val Thr Gly Ala Leu Glu	
365 370 375 380	

41	42
CTG TTG CAG CAG CTC TTC AGA ACG CCT CCA CCC GAG CTT CTG CAA ACC Leu Leu Gln Gln Leu Phe Arg Thr Pro Pro Pro Glu Leu Leu Gln Thr 385 390 395	1503
CTG ACC GCA GTC GGG GGC ATT GGG CAG CTC ACC GCT GCT AAG GAG GAG Leu Thr Ala Val Gly Gly Ile Gly Gln Leu Thr Ala Ala Lys Glu Glu 400 405 410	1551
TCT GGT GGC CGA AGC CGT AGT GGG AGT ATT GTG GAA CTT ATA GCT CGA Ser Gly Gly Arg Ser Arg Ser Gly Ser Ile Val Glu Leu Ile Ala Gly 415 420 425	1599
GGG GGT TCC TCA TGC AGC CCT GTC CTT TCA AGA AAA CAA AAA GGC AAA Gly Gly Ser Ser Cys Ser Pro Val Leu Ser Arg Lys Gln Lys Gly Lys 430 435 440	1647
GTG CTC TTA CGA GAA GAA GAA GCC TTG GAG GAT GAC TCT GAA TCG AGA Val Leu Leu Gly Glu Glu Glu Ala Leu Glu Asp Asp Ser Glu Ser Arg 445 450 455 460	1695
TCG GAT GTC AGC AGC TCT GCC TTA ACA GCC TCA GTG AAG GAT GAG ATC Ser Asp Val Ser Ser Ser Ala Leu Thr Ala Ser Val Lys Asp Glu Ile 465 470 475	1743
AGT GGA GAG CTG GCT GCT TCT TCA GGG GTT TCC ACT CCA GGG TCA GCA Ser Gly Glu Leu Ala Ala Ser Ser Gly Val Ser Thr Pro Gly Ser Ala 480 485 490	1791
CGT CAT GAC ATC ATC ACA GAA CAG CCA CGG TCA CAG CAC ACA CTG CAG Gly His Asp Ile Ile Thr Glu Gln Pro Arg Ser Gln His Thr Leu Gln 495 500 505	1839
CCG GAC TCA CTG GAT CTG GCC AGC TGT GAC TTG ACA AGC TCT GCC ACT Ala Asp Ser Leu Asp Leu Ala Ser Cys Asp Leu Thr Ser Ser Ala Thr 510 515 520	1887
CAT GCG GAT GAG GAG GAT ATC TTG AGC CAC AGC TCC AGC CAG GTC AGC Asp Gly Asp Glu Glu Asp Ile Leu Ser His Ser Ser Ser Gln Val Ser 525 530 535 540	1935
CCC GTC CCA TCT GAC CCT GCC ATG GAC CTG AAT GAT GGG ACC CAG CCC Ala Val Pro Ser Asp Pro Ala Met Asp Leu Asn Asp Gly Thr Gln Ala 545 550 555	1983
TCG TCG CCC ATC AGC GAC AGC TCC CAG ACC ACC ACC GAA GGG CCT GAT Ser Ser Pro Ile Ser Asp Ser Ser Gln Thr Thr Thr Glu Gly Pro Asp 560 565 570	2031
TCA GCT GTT ACC CCT TCA GAC AGT TCT GAA ATT GTG TTA GAC GGT ACC Ser Ala Val Thr Pro Ser Asp Ser Ser Glu Ile Val Leu Asp Gly Thr 575 580 585	2079
GAC AAC CAG TAT TTG GGC CTG CAG ATT GGA CAG CCC CAG GAT GAA GAT Asp Asn Gln Tyr Leu Gly Leu Gln Ile Gly Gln Pro Gln Asp Glu Asp 590 595 600	2127
GAG GAA GCC ACA GGT ATT CTT CCT GAT GAA GCC TCG GAG GCC TTC ACG Glu Glu Ala Thr Gly Ile Leu Pro Asp Glu Ala Ser Glu Ala Phe Arg 605 610 615 620	2175
AAC TCT TCC ATG GCC CTT CAA CAG GCA CAT TTA TTG AAA AAC ATG AGT Asn Ser Ser Met Ala Leu Gln Gln Ala His Leu Leu Lys Asn Met Ser 625 630 635	2223
CAC TGC AGG CAG CCT TCT GAC AGC AGT GTT GAT AAA TTT GTG TTG AGA His Cys Arg Gln Pro Ser Asp Ser Ser Val Asp Lys Phe Val Leu Arg	2271



43	640	645	650	44
	GAT GAA GCT ACT GAA CCG GGT GAT CAA GAA AAC AAG CCT TGC CGC ATC			2319
	Asp Glu Ala Thr Glu Pro Gly Asp Gln Glu Asn Lys Pro Cys Arg Ile			
	655	660	665	
	AAA GGT GAC ATT GGA CAG TCC ACT GAT GAT GAC TCT GCA CCT CTT GTC			2367
	Lys Gly Asp Ile Gly Gln Ser Thr Asp Asp Asp Ser Ala Pro Leu Val			
	670	675	680	
	CAT TCT GTC CGC CTT TTA TCT GCT TCG TTT TTG CTA ACA GGG GGA AAA			2415
	His Ser Val Arg Leu Leu Ser Ala Ser Phe Leu Leu Thr Gly Gly Lys			
	685	690	695	700
	AAT GTG CTG GTT CCG GAC ACG GAT GTG AGG GTC ACC GTG AAG GCC CTG			2463
	Asn Val Leu Val Pro Asp Arg Asp Val Arg Val Ser Val Lys Ala Leu			
	705	710	715	
	CCC CTC AGC TGT GTG GGA GCA GCT GTG GCC CTC CAC CCG GAA TCT TTC			2511
	Ala Leu Ser Cys Val Gly Ala Ala Val Ala Leu His Pro Glu Ser Phe			
	720	725	730	
	TTC ACC AAA CTC TAT AAA GTT CCT CTT GAC ACC ACG GAA TAC CCT GAG			2559
	Phe Ser Lys Leu Tyr Lys Val Pro Leu Asp Thr Thr Glu Tyr Pro Glu			
	735	740	745	
	GAA CAG TAT GTC TCA GAC ATC TTG AAC TAC ATC GAT CAT GGA GAC CCA			2607
	Glu Gln Tyr Val Ser Asp Ile Leu Asn Tyr Ile Asp His Gly Asp Pro			
	750	755	760	
	CAG GTT CGA GGA GCC ACT GCC ATT CTC TGT GGG ACC CTC ATC TGC TCC			2655
	Gln Val Arg Gly Ala Thr Ala Ile Leu Cys Gly Thr Leu Ile Cys Ser			
	765	770	775	780
	ATC CTC AGC AGG TCC CGC TTC CAC GTG GGA GAT TCG ATG GGC ACC ATT			2703
	Ile Leu Ser Arg Ser Arg Phe His Val Gly Asp Trp Met Gly Thr Ile			
	785	790	795	
	AGA ACC CTC ACA GGA AAT ACA TTT TCT TTG CCG GAT TGC ATT CCT TTG			2751
	Arg Thr Leu Thr Gly Asn Thr Phe Ser Leu Ala Asp Cys Ile Pro Leu			
	800	805	810	
	CTG CCG AAA ACA CTG AAG GAT GAG TCT TCT GTT ACT TGC AAG TTA GCT			2799
	Leu Arg Lys Thr Leu Lys Asp Glu Ser Ser Val Thr Cys Lys Leu Ala			
	815	820	825	
	TGT ACA GCT GTG AGG AAC TGT GTC ATG AGT CTC TCC AGC AGC AGC TAC			2847
	Cys Thr Ala Val Arg Asn Cys Val Met Ser Leu Cys Ser Ser Ser Tyr			
	830	835	840	
	AGT GAG TTA GGA CTG CAG CTG ATC ATC GAT GTG CTG ACT CTG AGG AAC			2895
	Ser Glu Leu Gly Leu Gln Leu Ile Ile Asp Val Leu Thr Leu Arg Asn			
	845	850	855	860
	AGT TCC TAT TCG CTG GTG ACG ACA GAG CTT CTG GAA ACC CTT GCA GAG			2943
	Ser Ser Tyr Trp Leu Val Arg Thr Glu Leu Leu Glu Thr Leu Ala Glu			
	865	870	875	
	ATT GAC TTC AGC CTG GTG ACC TTT TTG GAG GCA AAA GCA GAA AAC TTA			2991
	Ile Asp Phe Arg Leu Val Ser Phe Leu Glu Ala Lys Ala Glu Asn Leu			
	880	885	890	
	CAC AGA GGG GCT CAT CAT TAT ACA GGG CTT TTA AAA CTG CAA GAA CGA			3039
	His Arg Gly Ala His His Tyr Thr Gly Leu Leu Lys Leu Gln Glu Arg			
	895	900	905	
	GTG CTC AAT AAT GTT GTC ATC CAT TTG CTT CGA GAT GAA GAC CCC AGG			3087

45	46
Val Leu Asn Asn Val Val Ile His Leu Leu Gly Asp Glu Asp Pro Arg	
910 915 920	
GTG CGA CAT GTT GCC GCA GCA TCA CTA ATT AGG CTT GTC CCA AAG CTG	3135
Val Arg His Val Ala Ala Ala Ser Leu Ile Arg Leu Val Pro Lys Leu	
925 930 935 940	
TTT TAT AAA TGT GAC CAA GGA CAA GCT GAT CCA GTA GTG GCC GTG CCA	3183
Phe Tyr Lys Cys Asp Gln Gly Gln Ala Asp Pro Val Val Ala Val Ala	
945 950 955	
AGA GAT CAA AGC AGT GTT TAC CTG AAA CTT CTC ATG CAT GAG ACG CAG	3231
Arg Asp Gln Ser Ser Val Tyr Leu Lys Leu Leu Met His Glu Thr Gln	
960 965 970	
CCT CCA TCT CAT TTC TCC GTC AGC ACA ATA ACC AGA ATA TAT AGA GGC	3279
Pro Pro Ser His Phe Ser Val Ser Thr Ile Thr Arg Ile Tyr Arg Gly	
975 980 985	
TAT AAC CTA CTA CCA AGC ATA ACA GAC GTC ACT ATG GAA AAT AAC CTT	3327
Tyr Asn Leu Leu Pro Ser Ile Thr Asp Val Thr Met Glu Asn Asn Leu	
990 995 1000	
TCA AGA GTT ATT GCA GCA GTT TCT CAT GAA CTA ATC ACA TCA ACC ACC	3375
Ser Arg Val Ile Ala Ala Val Ser His Glu Leu Ile Thr Ser Thr Thr	
1005 1010 1015 1020	
AGA GCA CTC ACA TTT GGA TGC TGT GAA GCT TTG TGT CTT CTT TCC ACT	3423
Arg Ala Leu Thr Phe Gly Cys Cys Glu Ala Leu Cys Leu Leu Ser Thr	
1025 1030 1035	
GCC TTC CCA GTT TGC ATT TCG AGT TTA GGT TGG CAC TGT GGA GTG CCT	3471
Ala Phe Pro Val Cys Ile Trp Ser Leu Gly Trp His Cys Gly Val Pro	
1040 1045 1050	
CCA CTG AGT GCC TCA GAT GAG TCT AGG AAG AGC TGT ACC GTT GGG ATG	3519
Pro Leu Ser Ala Ser Asp Glu Ser Arg Lys Ser Cys Thr Val Gly Met	
1055 1060 1065	
GCC ACA ATG ATT CTG ACC CTG CTC TCG TCA GCT TCG TTC CCA TTG GAT	3567
Ala Thr Met Ile Leu Thr Leu Leu Ser Ser Ala Trp Phe Pro Leu Asp	
1070 1075 1080	
CTC TCA GCC CAT CAA GAT GCT TTG ATT TTG GCC GGA AAC TTG CTT GCA	3615
Leu Ser Ala His Gln Asp Ala Leu Ile Leu Ala Gly Asn Leu Leu Ala	
1085 1090 1095 1100	
GCC AGT GCT CCC AAA TCT CTG AGA AGT TCA TGG GCC TCT GAA GAA GAA	3663
Ala Ser Ala Pro Lys Ser Leu Arg Ser Ser Trp Ala Ser Glu Glu Glu	
1105 1110 1115	
GCC AAC CCA GCA GCC ACC AAG CAA GAG GAG GTC TGG CCA GCC CTG GGG	3711
Ala Asn Pro Ala Ala Thr Lys Gln Glu Glu Val Trp Pro Ala Leu Gly	
1120 1125 1130	
GAC CGG GCC CTG GTG CCC ATG GTG GAG CAG CTC TTC TCT CAC CTG CTG	3759
Asp Arg Ala Leu Val Pro Met Val Glu Gln Leu Phe Ser His Leu Leu	
1135 1140 1145	
AAG GTG ATT AAC ATT TGT GCC CAC GTC CTG GAT GAC GTG GCT CCT CGA	3807
Lys Val Ile Asn Ile Cys Ala His Val Leu Asp Asp Val Ala Pro Gly	
1150 1155 1160	
CCC GCA ATA AAG GCA GCC TTG CCT TCT CTA ACA AAC CCC CCT TCT CTA	3855
Pro Ala Ile Lys Ala Ala Leu Pro Ser Leu Thr Asn Pro Pro Ser Leu	
1165 1170 1175 1180	



47	48
AGT CCC ATC CGA CGA AAG GCG AAG GAG AAA GAA CCA GGA GAA CAA CCA Ser Pro Ile Arg Arg Lys Gly Lys Glu Lys Glu Pro Gly Glu Gln Ala 1185 1190 1195	3903
TCT GTA CCG TTG AGT CCC AAG AAA GGC AGT GAG GCC AGT GCA GCT TCT Ser Val Pro Leu Ser Pro Lys Lys Gly Ser Glu Ala Ser Ala Ala Ser 1200 1205 1210	3951
AGA CAA TCT GAT ACC TCA GGT CCT GTT ACA ACA AGT AAA TCC TCA TCA Arg Gln Ser Asp Thr Ser Gly Pro Val Thr Thr Ser Lys Ser Ser Ser 1215 1220 1225	3999
CTG GCG AGT TTC TAT CAT CTT CCT TCA TAC CTC AGA CTG CAT GAT GTC Leu Gly Ser Phe Tyr His Leu Pro Ser Tyr Leu Arg Leu His Asp Val 1230 1235 1240	4047
CTG AAA GCT ACA CAC GCT AAC TAC AAG GTC ACG CTG GAT CTT CAG AAC Leu Lys Ala Thr His Ala Asn Tyr Lys Val Thr Leu Asp Leu Gln Asn 1245 1250 1255 1260	4095
AGC ACG GAA AAG TTT CGA GCG TTT CTC CGC TCA GCC TTG GAT GTT CTT Ser Thr Glu Lys Phe Gly Gly Phe Leu Arg Ser Ala Leu Asp Val Leu 1265 1270 1275	4143
TCT CAG ATA CTA GAG CTG GCC ACA CTG CAG GAC ATT GGG AAG TGT GTT Ser Gln Ile Leu Glu Leu Ala Thr Leu Gln Asp Ile Gly Lys Cys Val 1280 1285 1290	4191
GAA GAG ATC CTA GGA TAC CTG AAA TCC TGC TTT AGT CGA GAA CCA ATG Glu Glu Ile Leu Gly Tyr Leu Lys Ser Cys Phe Ser Arg Glu Pro Met 1295 1300 1305	4239
ATG GCA ACT GTT TGT GTT CAA CAA TTG TTG AAG ACT CTC TTT GGC ACA Met Ala Thr Val Cys Val Gln Gln Leu Leu Lys Thr Leu Phe Gly Thr 1310 1315 1320	4287
AAC TTG GCC TCC CAG TTT GAT GGC TTA TCT TCC AAC CCC AGC AAG TCA Asn Leu Ala Ser Gln Phe Asp Gly Leu Ser Ser Asn Pro Ser Lys Ser 1325 1330 1335 1340	4335
CAA GCC CGA GCA CAG CGC CTT GGC TCC TCC AGT GTG AGG CCA GCC TTG Gln Gly Arg Ala Gln Arg Leu Gly Ser Ser Ser Val Arg Pro Gly Leu 1345 1350 1355	4383
TAC CAC TAC TGC TTC ATG GCC CCG TAC ACC CAC TTC ACC CAG GCC CTC Tyr His Tyr Cys Phe Met Ala Pro Tyr Thr His Phe Thr Gln Ala Leu 1360 1365 1370	4431
GCT GAC GCC AGC CTG AGG AAC ATG GTG CAG GCG GAG CAG GAG AAC GAC Ala Asp Ala Ser Leu Arg Asn Met Val Gln Ala Glu Gln Glu Asn Asp 1375 1380 1385	4479
ACC TCG GGA TCG TTT GAT GTC CTC CAG AAA GTG TCT ACC CAG TTG AAG Thr Ser Gly Trp Phe Asp Val Leu Gln Lys Val Ser Thr Gln Leu Lys 1390 1395 1400	4527
ACA AAC CTC ACG AGT GTC ACA AAG AAC CGT GCA GAT AAG AAT GCT ATT Thr Asn Leu Thr Ser Val Thr Lys Asn Arg Ala Asp Lys Asn Ala Ile 1405 1410 1415 1420	4575
CAT AAT CAC ATT CGT TTG TTT GAA CCT CTT GTT ATA AAA GCT TTA AAA His Asn His Ile Arg Leu Phe Glu Pro Leu Val Ile Lys Ala Leu Lys 1425 1430 1435	4623
CAG TAC ACG ACT ACA ACA TGT GTG CAG TTA CAG AAG CAG GTT TTA GAT Gln Tyr Thr Thr Thr Thr Cys Val Gln Leu Gln Lys Gln Val Leu Asp	4671

49		50
	1440	1445
TTG CTG GCG CAG CTG GTT CAG TTA CGG GTT AAT TAC TGT CTT CTG GAT		1450
Leu Leu Ala Gln Leu Val Gln Leu Arg Val Asn Tyr Cys Leu Leu Asp		
1455	1460	1465
TCA GAT CAG GTG TTT ATT GCC TTT GTA TTG AAA CAG TTT GAA TAC ATT		
Ser Asp Gln Val Phe Ile Gly Phe Val Leu Lys Gln Phe Glu Tyr Ile		
1470	1475	1480
GAA GTG GCG CAG TTC AGG GAA TCA GAG GCA ATC ATT CCA AAC ATC TTT		
Glu Val Gly Gln Phe Arg Glu Ser Glu Ala Ile Ile Pro Asn Ile Phe		
1485	1490	1495
TTC TTC TTG GTA TTA CTA TCT TAT GAA CGC TAT CAT TCA AAA CAG ATC		
Phe Phe Leu Val Leu Leu Ser Tyr Glu Arg Tyr His Ser Lys Gln Ile		
1505	1510	1515
ATT GGA ATT CCT AAA ATC ATT CAG CTC TGT GAT GCC ATC ATG GCC AGT		
Ile Gly Ile Pro Lys Ile Ile Gln Leu Cys Asp Gly Ile Met Ala Ser		
1520	1525	1530
GGA AGG AAG GCT GTG ACA CAT GCC ATA CCG GCT CTG CAG CCC ATA GTC		
Gly Arg Lys Ala Val Thr His Ala Ile Pro Ala Leu Gln Pro Ile Val		
1535	1540	1545
CAC GAC CTC TTT GTA TTA AGA GGA ACA AAT AAA GCT GAT GCA GGA AAA		
His Asp Leu Phe Val Leu Arg Gly Thr Asn Lys Ala Asp Ala Gly Lys		
1550	1555	1560
GAG CTT GAA ACC CAA AAA GAG GTG GTG GTG TCA ATG TTA CTG AGA CTC		
Glu Leu Glu Thr Gln Lys Glu Val Val Val Ser Met Leu Leu Arg Leu		
1565	1570	1575
ATC CAG TAC CAT CAG GTG TTG GAG ATG TTC ATT CTT GTC CTG CAG CAG		
Ile Gln Tyr His Gln Val Leu Glu Met Phe Ile Leu Val Leu Gln Gln		
1585	1590	1595
TGC CAC AAG GAG AAT GAA GAC AAG TGG AAG CGA CTG TCT CGA CAG ATA		
Cys His Lys Glu Asn Glu Asp Lys Trp Lys Arg Leu Ser Arg Gln Ile		
1600	1605	1610
GCT GAC ATC ATC CTC CCA ATG TTA GCC AAA CAG CAG ATG CAC ATT GAC		
Ala Asp Ile Ile Leu Pro Met Leu Ala Lys Gln Gln Met His Ile Asp		
1615	1620	1625
TCT CAT GAA GCC CTT CGA GTG TTA AAT ACA TTA TTT GAG ATT TTG GCC		
Ser His Glu Ala Leu Gly Val Leu Asn Thr Leu Phe Glu Ile Leu Ala		
1630	1635	1640
CCT TCC TCC CTC CGT CCG GTA GAC ATG CTT TTA CCG AGT ATG TTC GTC		
Pro Ser Ser Leu Arg Pro Val Asp Met Leu Leu Arg Ser Met Phe Val		
1645	1650	1655
ACT CCA AAC ACA ATG CCG TCC GTG AGC ACT GTT CAA CTG TGG ATA TCG		
Thr Pro Asn Thr Met Ala Ser Val Ser Thr Val Gln Leu Trp Ile Ser		
1665	1670	1675
GGA ATT CTG GCC ATT TTG ACG GTT CTG ATT TCC CAG TCA ACT GAA GAT		
Gly Ile Leu Ala Ile Leu Arg Val Leu Ile Ser Gln Ser Thr Glu Asp		
1680	1685	1690
ATT GTT CTT TCT CGT ATT CAG GAG CTC TCC TTC TCT CCG TAT TTA ATC		
Ile Val Leu Ser Arg Ile Gln Glu Leu Ser Phe Ser Pro Tyr Leu Ile		
1695	1700	1705
TCC TGT ACA GTA ATT AAT ACG TTA AGA GAT GGG GAC AGT ACT TCA ACG		



51	52
Ser Cys Thr Val Ile Asn Arg Leu Arg Asp Gly Asp Ser Thr Ser Thr	
1710 1715 1720	
CTA GAA GAA CAC AGT GAA GCG AAA CAA ATA AAG AAT TTG CCA GAA GAA	5535
Leu Glu Glu His Ser Glu Gly Lys Gln Ile Lys Asn Leu Pro Glu Glu	
1725 1730 1735 1740	
ACA TTT TCA AGG TTT CTA TTA CAA CTG GTT GGT ATT CTT TTA GAA GAC	5583
Thr Phe Ser Arg Phe Leu Leu Gln Leu Val Gly Ile Leu Leu Glu Asp	
1745 1750 1755	
ATT GTT ACA AAA CAG CTG AAG GTG GAA ATG AGT GAG CAG CAA CAT ACT	5631
Ile Val Thr Lys Gln Leu Lys Val Glu Met Ser Glu Gln Gln His Thr	
1760 1765 1770	
TTC TAT TGC CAG GAA CTA GGC ACA CTG CTA ATG TGT CTG ATC CAC ATC	5679
Phe Tyr Cys Gln Glu Leu Gly Thr Leu Leu Met Cys Leu Ile His Ile	
1775 1780 1785	
TTC AAG TCT GGA ATG TTC CCG AGA ATC ACA GCA GCT GCC ACT AGG CTG	5727
Phe Lys Ser Gly Met Phe Arg Arg Ile Thr Ala Ala Ala Thr Arg Leu	
1790 1795 1800	
TTC CCC AGT GAT GGC TGT GGC GGC AGT TTC TAC ACC CTG GAC AGC TTG	5775
Phe Arg Ser Asp Gly Cys Gly Gly Ser Phe Tyr Thr Leu Asp Ser Leu	
1805 1810 1815 1820	
AAC TTG CCG GCT CGT TCC ATG ATC ACC ACC CAC CCG GCC CTG GTG CTG	5823
Asn Leu Arg Ala Arg Ser Met Ile Thr Thr His Pro Ala Leu Val Leu	
1825 1830 1835	
CTC TGG TGT CAG ATA CTG CTG CTT GTC AAC CAC ACC GAC TAC CGC TGG	5871
Leu Trp Cys Gln Ile Leu Leu Leu Val Asn His Thr Asp Tyr Arg Trp	
1840 1845 1850	
TGG GCA GAA GTG CAG CAG ACC CCG AAA AGA CAC AGT CTG TCC AGC ACA	5919
Trp Ala Glu Val Gln Gln Thr Pro Lys Arg His Ser Leu Ser Ser Thr	
1855 1860 1865	
AAG TTA CTT AGT CCC CAG ATG TCT CGA GAA GAG GAG GAT TCT GAC TTG	5967
Lys Leu Leu Ser Pro Gln Met Ser Gly Glu Glu Glu Asp Ser Asp Leu	
1870 1875 1880	
GCA GCC AAA CTT GGA ATG TGC AAT AGA GAA ATA GTA CGA AGA GCG GCT	6015
Ala Ala Lys Leu Gly Met Cys Asn Arg Glu Ile Val Arg Arg Gly Ala	
1885 1890 1895 1900	
CTC ATT CTC TTC TGT GAT TAT GTC TGT CAG AAC CTC CAT GAC TCC GAG	6063
Leu Ile Leu Phe Cys Asp Tyr Val Cys Gln Asn Leu His Asp Ser Glu	
1905 1910 1915	
CAC TTA ACG TGG CTC ATT GTA AAT CAC ATT CAA GAT CTG ATC AGC CTT	6111
His Leu Thr Trp Leu Ile Val Asn His Ile Gln Asp Leu Ile Ser Leu	
1920 1925 1930	
TCC CAC GAG CCT CCA GTA CAG GAC TTC ATC AGT GCC GTT CAT CGG AAC	6159
Ser His Glu Pro Pro Val Gln Asp Phe Ile Ser Ala Val His Arg Asn	
1935 1940 1945	
TCT GCT GCC AGC GGC CTG TTC ATC CAG GCA ATT CAG TCT CGT TGT GAA	6207
Ser Ala Ala Ser Gly Leu Phe Ile Gln Ala Ile Gln Ser Arg Cys Glu	
1950 1955 1960	
AAC CTT TCA ACT CCA ACC ATG CTG AAG AAA ACT CTT CAG TGC TTG GAG	6255
Asn Leu Ser Thr Pro Thr Met Leu Lys Lys Thr Leu Gln Cys Leu Glu	
1965 1970 1975 1980	

53	54
CGG ATC CAT CTC AGC CAG TCG GGA GCT GTG CTC ACG CTG TAT GTG GAC Gly Ile His Leu Ser Gln Ser Gly Ala Val Leu Thr Leu Tyr Val Asp	6303
1985 1990 1995	
AGG CTT CTG TGC ACC CCT TTC CGT GTG CTG GCT CGC ATG GTC GAC ATC Arg Leu Leu Cys Thr Pro Phe Arg Val Leu Ala Arg Met Val Asp Ile	6351
2000 2005 2010	
CTT GCT TGT CGC CGG GTA GAA ATG CTT CTG GCT GCA AAT TTA CAG AGC Leu Ala Cys Arg Arg Val Glu Met Leu Leu Ala Ala Asn Leu Gln Ser	6399
2015 2020 2025	
AGC ATG GCC CAG TTG CCA ATG GAA GAA CTC AAC AGA ATC CAG GAA TAC Ser Met Ala Gln Leu Pro Met Glu Glu Leu Asn Arg Ile Gln Glu Tyr	6447
2030 2035 2040	
CTT CAG AGC AGC GGG CTC GCT CAG AGA CAC CAA AGG CTC TAT TCC CTG Leu Gln Ser Ser Gly Leu Ala Gln Arg His Gln Arg Leu Tyr Ser Leu	6495
2045 2050 2055 2060	
CTG GAC AGG TTT CGT CTC TCC ACC ATG CAA GAC TCA CTT AGT CCC TCT Leu Asp Arg Phe Arg Leu Ser Thr Met Gln Asp Ser Leu Ser Pro Ser	6543
2065 2070 2075	
CCT CCA GTC TCT TCC CAC CCG CTG GAC GGG GAT GGG CAC GTG TCA CTG Pro Pro Val Ser Ser His Pro Leu Asp Gly Asp Gly His Val Ser Leu	6591
2080 2085 2090	
GAA ACA GTG AGT CCG GAC AAA GAC TGG TAC GTT CAT CTT GTC AAA TCC Glu Thr Val Ser Pro Asp Lys Asp Trp Tyr Val His Leu Val Lys Ser	6639
2095 2100 2105	
CAG TGT TGG ACC AGG TCA GAT TCT GCA CTG CTG GAA GGT GCA GAG CTG Gln Cys Trp Thr Arg Ser Asp Ser Ala Leu Leu Glu Gly Ala Glu Leu	6687
2110 2115 2120	
GTG AAT CCG ATT CCT GCT GAA GAT ATG AAT GCC TTC ATG ATG AAC TCG Val Asn Arg Ile Pro Ala Glu Asp Met Asn Ala Phe Met Met Asn Ser	6735
2125 2130 2135 2140	
CAG TTC AAC CTA AGC CTG CTA GCT CCA TGC TTA AGC CTA CGG ATG AGT Glu Phe Asn Leu Ser Leu Leu Ala Pro Cys Leu Ser Leu Gly Met Ser	6783
2145 2150 2155	
GAA ATT TCT GGT GGC CAG AAG AGT GCC CTT TTT GAA GCA GCC CGT GAG Glu Ile Ser Gly Gly Gln Lys Ser Ala Leu Phe Glu Ala Ala Arg Glu	6831
2160 2165 2170	
GTG ACT CTG GCC CGT GTG AGC GGC ACC GTG CAG CAG CTC CCT GCT GTC Val Thr Leu Ala Arg Val Ser Gly Thr Val Gln Gln Leu Pro Ala Val	6879
2175 2180 2185	
CAT CAT GTC TTC CAG CCC GAG CTG CCT GCA GAG CCG GCG GCC TAC TGG His His Val Phe Gln Pro Glu Leu Pro Ala Glu Pro Ala Ala Tyr Trp	6927
2190 2195 2200	
AGC AAG TTG AAT GAT CTG TTT GCG GAT GCT GCA CTG TAT CAG TCC CTG Ser Lys Leu Asn Asp Leu Phe Gly Asp Ala Ala Leu Tyr Gln Ser Leu	6975
2205 2210 2215 2220	
CCC ACT CTG GCC CGG GCC CTG GCA CAG TAC CTG GTG GTG GTC TCC AAA Pro Thr Leu Ala Arg Ala Leu Ala Gln Tyr Leu Val Val Val Ser Lys	7023
2225 2230 2235	
CTG CCC AGT CAT TTG CAC CTT CCT CCT GAG AAA GAG AAG GAC ATT GTG Leu Pro Ser His Leu His Leu Pro Pro Glu Lys Glu Lys Asp Ile Val	7071

55			56
	2240	2245	2250
	AAA TTC GTG GTG GCA ACC CTT GAG GCC CTG TCC TGG CAT TTG ATC CAT		7119
	Lys Phe Val Val Ala Thr Leu Glu Ala Leu Ser Trp His Leu Ile His		
	2255	2260	2265
	GAG CAG ATC CCG CTG AGT CTG GAT CTC CAG GCA GCG CTG GAC TGC TGC		7167
	Glu Gln Ile Pro Leu Ser Leu Asp Leu Gln Ala Gly Leu Asp Cys Cys		
	2270	2275	2280
	TGC CTG GCC CTG CAG CTG CCT GGC CTC TGG AGC GTG GTC TCC TCC ACA		7215
	Cys Leu Ala Leu Gln Leu Pro Gly Leu Trp Ser Val Val Ser Ser Thr		
	2285	2290	2295
	GAG TTT GTG ACC CAC GCC TCC TCC CTC ATC TAC TGT GTG CAC TTC ATC		7263
	Glu Phe Val Thr His Ala Cys Ser Leu Ile Tyr Cys Val His Phe Ile		
	2305	2310	2315
	CTG GAG GCC GTT GCA GTG CAG CCT GGA GAG CAG CTT CTT AGT CCA GAA		7311
	Leu Glu Ala Val Ala Val Gln Pro Gly Glu Gln Leu Leu Ser Pro Glu		
	2320	2325	2330
	AGA ACG ACA AAT ACC CCA AAA GCC ATC AGC GAG GAG GAG GAG GAA GTA		7359
	Arg Arg Thr Asn Thr Pro Lys Ala Ile Ser Glu Glu Glu Glu Glu Val		
	2335	2340	2345
	GAT CCA AAC ACA CAG AAT CCT AAG TAT ATC ACT GCA CCC TGT GAG ATG		7407
	Asp Pro Asn Thr Gln Asn Pro Lys Tyr Ile Thr Ala Ala Cys Glu Met		
	2350	2355	2360
	GTG GCA GAA ATG GTG GAG TCT CTG CAG TCG GTG TTG GCC TTG GGT CAT		7455
	Val Ala Glu Met Val Glu Ser Leu Gln Ser Val Leu Ala Leu Gly His		
	2365	2370	2375
	AAA ACG AAT AGC GCC GTG CCG GCG TTT CTC ACG CCA TTG CTC AGG AAC		7503
	Lys Arg Asn Ser Gly Val Pro Ala Phe Leu Thr Pro Leu Leu Arg Asn		
	2385	2390	2395
	ATC ATC ATC AGC CTG GCC CCC CTG CCC CTT GTC AAC AGC TAC ACA CGT		7551
	Ile Ile Ile Ser Leu Ala Arg Leu Pro Leu Val Asn Ser Tyr Thr Arg		
	2400	2405	2410
	GTG CCC CCA CTG GTG TGG AAG CTT GGA TGG TCA CCC AAA CCG GGA GCG		7599
	Val Pro Pro Leu Val Trp Lys Leu Gly Trp Ser Pro Lys Pro Gly Gly		
	2415	2420	2425
	GAT TTT GGC ACA GCA TTC CCT GAG ATC CCC GTG GAG TTC CTC CAG GAA		7647
	Asp Phe Gly Thr Ala Phe Pro Glu Ile Pro Val Glu Phe Leu Gln Glu		
	2430	2435	2440
	AAG GAA GTC TTT AAG GAG TTC ATC TAC CGC ATC AAC ACA CTA GGC TGG		7695
	Lys Glu Val Phe Lys Glu Phe Ile Tyr Arg Ile Asn Thr Leu Gly Trp		
	2445	2450	2455
	ACC AGT CGT ACT CAG TTT GAA GAA ACT TGG GCC ACC CTC CTT GGT GTC		7743
	Thr Ser Arg Thr Gln Phe Glu Glu Thr Trp Ala Thr Leu Leu Gly Val		
	2465	2470	2475
	CTG GTG ACG CAG CCC CTC GTG ATG GAG CAG GAG GAG AGC CCA CCA GAA		7791
	Leu Val Thr Gln Pro Leu Val Met Glu Gln Glu Glu Ser Pro Pro Glu		
	2480	2485	2490
	GAA GAC ACA GAG AGG ACC CAG ATC AAC GTC CTG GCC GTG CAG GCC ATC		7839
	Glu Asp Thr Glu Arg Thr Gln Ile Asn Val Leu Ala Val Gln Ala Ile		
	2495	2500	2505
	ACC TCA CTG GTG CTC AGT GCA ATG ACT GTG CCT GTG GCC GCC AAC CCA		7887



57	58
Thr Ser Leu Val Leu Ser Ala Met Thr Val Pro Val Ala Gly Asn Pro	
2510 2515 2520	
CCT GTA AGC TGC TTG GAG CAG CAG CCC CGG AAC AAG CCT CTG AAA CCT	7935
Ala Val Ser Cys Leu Glu Gln Gln Pro Arg Asn Lys Pro Leu Lys Ala	
2525 2530 2535 2540	
CTC GAC ACC AGG TTT GGG AGG AAG CTG AGC ATT ATC AGA GGG ATT GTG	7983
Leu Asp Thr Arg Phe Gly Arg Lys Leu Ser Ile Ile Arg Gly Ile Val	
2545 2550 2555	
GAG CAA GAG ATT CAA GCA ATG GTT TCA AAG AGA GAG AAT ATT GCC ACC	8031
Glu Gln Glu Ile Gln Ala Met Val Ser Lys Arg Glu Asn Ile Ala Thr	
2560 2565 2570	
CAT CAT TTA TAT CAG GCA TGG GAT CCT GTC CCT TCT CTG TCT CCG GCT	8079
His His Leu Tyr Gln Ala Trp Asp Pro Val Pro Ser Leu Ser Pro Ala	
2575 2580 2585	
ACT ACA GGT GCC CTC ATC AGC CAC GAG AAG CTG CTG CTA CAG ATC AAC	8127
Thr Thr Gly Ala Leu Ile Ser His Glu Lys Leu Leu Leu Gln Ile Asn	
2590 2595 2600	
CCC GAG CCG GAG CTG GGG AGC ATG AGC TAC AAA CTC GGC CAG GTG TCC	8175
Pro Glu Arg Glu Leu Gly Ser Met Ser Tyr Lys Leu Gly Gln Val Ser	
2605 2610 2615 2620	
ATA CAC TCC GTG TGG CTG GGG AAC AGC ATC ACA CCC CTG AGG GAG GAG	8223
Ile His Ser Val Trp Leu Gly Asn Ser Ile Thr Pro Leu Arg Glu Glu	
2625 2630 2635	
GAA TCG GAC GAG GAA GAG GAG GAG GAG GCC GAC GCC CCT GCA CCT TCG	8271
Glu Trp Asp Glu Glu Glu Glu Glu Glu Ala Asp Ala Pro Ala Pro Ser	
2640 2645 2650	
TCA CCA CCC ACG TCT CCA GTC AAC TCC AGG AAA CAC CGG GCT GGA GTT	8319
Ser Pro Pro Thr Ser Pro Val Asn Ser Arg Lys His Arg Ala Gly Val	
2655 2660 2665	
GAC ATC CAC TCC TGT TCG CAG TTT TTG CTT GAG TTG TAC AGC CGC TGG	8367
Asp Ile His Ser Cys Ser Gln Phe Leu Leu Glu Leu Tyr Ser Arg Trp	
2670 2675 2680	
ATC CTG CCG TCC AGC TCA GCC AGG AGG ACC CCG GCC ATC CTG ATC AGT	8415
Ile Leu Pro Ser Ser Ser Ala Arg Arg Thr Pro Ala Ile Leu Ile Ser	
2685 2690 2695 2700	
GAG GTG GTC AGA TCC CTT CTA GTG GTC TCA GAC TTG TTC ACC GAG CGC	8463
Glu Val Val Arg Ser Leu Leu Val Val Ser Asp Leu Phe Thr Glu Arg	
2705 2710 2715	
AAC CAG TTT GAG CTG ATG TAT GTG ACG CTG ACA GAA CTG CGA AGG GTG	8511
Asn Gln Phe Glu Leu Met Tyr Val Thr Leu Thr Glu Leu Arg Arg Val	
2720 2725 2730	
CAC CCT TCA GAA GAC GAG ATC CTC GCT CAG TAC CTG GTG CCT GCC ACC	8559
His Pro Ser Glu Asp Glu Ile Leu Ala Gln Tyr Leu Val Pro Ala Thr	
2735 2740 2745	
TGC AAG GCA GCT GCC GTC CTT GGG ATG GAC AAG GCC GTG GCG GAG CCT	8607
Cys Lys Ala Ala Ala Val Leu Gly Met Asp Lys Ala Val Ala Glu Pro	
2750 2755 2760	
GTC ACC CGC CTG CTG GAG ACC ACG CTC AGG ACC ACC CAC CTG CCC AGC	8655
Val Ser Arg Leu Leu Glu Ser Thr Leu Arg Ser Ser His Leu Pro Ser	
2765 2770 2775 2780	

59	60
AGG GTT GGA GCC CTG CAC GGC ATC CTC TAT GTG CTG GAG TGC GAC CTG	8703
Arg Val Gly Ala Leu His Gly Ile Leu Tyr Val Leu Glu Cys Asp Leu	
2785 2790 2795	
CTG GAC GAC ACT GCC AAG CAG CTC ATC CCG GTC ATC AGC GAC TAT CTC	8751
Leu Asp Asp Thr Ala Lys Gln Leu Ile Pro Val Ile Ser Asp Tyr Leu	
2800 2805 2810	
CTC TCC AAC CTG AAA GGG ATC GCC CAC TGC GTG AAC ATT CAC AGC CAG	8799
Leu Ser Asn Leu Lys Gly Ile Ala His Cys Val Asn Ile His Ser Gln	
2815 2820 2825	
CAG CAC GTA CTG GTC ATG TGT GCC ACT GCG TTT TAC CTC ATT GAG AAC	8847
Gln His Val Leu Val Met Cys Ala Thr Ala Phe Tyr Leu Ile Glu Asn	
2830 2835 2840	
TAT CCT CTG GAC GTA GGG CCG GAA TTT TCA GCA TCA ATA ATA CAG ATG	8895
Tyr Pro Leu Asp Val Gly Pro Glu Phe Ser Ala Ser Ile Ile Gln Met	
2845 2850 2855 2860	
TGT GCG GTG ATG CTG TCT GGA AGT GAG GAG TCC ACC CCC TCC ATC ATT	8943
Cys Gly Val Met Leu Ser Gly Ser Glu Glu Ser Thr Pro Ser Ile Ile	
2865 2870 2875	
TAC CAC TGT GCC CTC AGA GGC CTG GAG CGC CTC CTG CTC TCT GAG CAG	8991
Tyr His Cys Ala Leu Arg Gly Leu Glu Arg Leu Leu Leu Ser Glu Gln	
2880 2885 2890	
CTC TCC CCG CTG GAT CCA GAA TCG CTG GTC AAG CTG AGT GTG GAC AGA	9039
Leu Ser Arg Leu Asp Ala Glu Ser Leu Val Lys Leu Ser Val Asp Arg	
2895 2900 2905	
GTG AAC GTG CAC AGC CCG CAC CCG GCC ATG CCG GCT CTG GGC CTG ATG	9087
Val Asn Val His Ser Pro His Arg Ala Met Ala Ala Leu Gly Leu Met	
2910 2915 2920	
CTC ACC TGC ATG TAC ACA GGA AAG GAG AAA GTC AGT CCG GGT AGA ACT	9135
Leu Thr Cys Met Tyr Thr Gly Lys Glu Lys Val Ser Pro Gly Arg Thr	
2925 2930 2935 2940	
TCA GAC CCT AAT CCT CCA GCC CCC GAC AGC GAG TCA GTG ATT GTT CCT	9183
Ser Asp Pro Asn Pro Ala Ala Pro Asp Ser Glu Ser Val Ile Val Ala	
2945 2950 2955	
ATG GAG CCG GTA TCT GTT CTT TTT GAT AGG ATC ACG AAA GGC TTT CCT	9231
Met Glu Arg Val Ser Val Leu Phe Asp Arg Ile Arg Lys Gly Phe Pro	
2960 2965 2970	
TGT GAA GCC AGA GTG GTG GCC AGG ATC CTG CCC CAG TTT CTA GAC GAC	9279
Cys Glu Ala Arg Val Val Ala Arg Ile Leu Pro Gln Phe Leu Asp Asp	
2975 2980 2985	
TTC TTC CCA CCC CAG GAC ATC ATG AAC AAA GTC ATC CGA GAG TTT CTG	9327
Phe Phe Pro Pro Gln Asp Ile Met Asn Lys Val Ile Gly Glu Phe Leu	
2990 2995 3000	
TCC AAC CAG CAG CCA TAC CCC CAG TTC ATG GCC ACC GTG GTG TAT AAG	9375
Ser Asn Gln Gln Pro Tyr Pro Gln Phe Met Ala Thr Val Val Tyr Lys	
3005 3010 3015 3020	
GTG TTT CAG ACT CTG CAC ACC ACC GGG CAG TCG TCC ATG GTC CCG GAC	9423
Val Phe Gln Thr Leu His Ser Thr Gly Gln Ser Ser Met Val Arg Asp	
3025 3030 3035	
TGG GTC ATG CTG TCC CTC TCC AAC TTC ACG CAG ACG GCC CCG GTC GCC	9471
Trp Val Met Leu Ser Leu Ser Asn Phe Thr Gln Arg Ala Pro Val Ala	

61			62
	3040	3045	3050
	ATG GCC ACG TGG AGC CTC TCC TGC TTC TTT GTC AGC GCG TCC ACC AGC		9519
	Met Ala Thr Trp Ser Leu Ser Cys Phe Phe Val Ser Ala Ser Thr Ser		
	3055	3060	3065
	CCG TCG GTC GCG GCG ATC CTC CCA CAT GTC ATC AGC AGG ATG GGC AAG		9567
	Pro Trp Val Ala Ala Ile Leu Pro His Val Ile Ser Arg Met Gly Lys		
	3070	3075	3080
	CTG GAG CAG GTG GAC GTG AAC CTT TTC TGC CTG GTC GCC ACA GAC TTC		9615
	Leu Glu Gln Val Asp Val Asn Leu Phe Cys Leu Val Ala Thr Asp Phe		
	3085	3090	3095
	TAC AGA CAC CAG ATA GAG GAG GAG CTC GAC CGC AGG GCC TTC CAG TCT		9663
	Tyr Arg His Gln Ile Glu Glu Glu Leu Asp Arg Arg Ala Phe Gln Ser		
	3105	3110	3115
	GTG CTT GAG GTG GTT GCA GCC CCA GGA AGC CCA TAT CAC CGG CTG CTG		9711
	Val Leu Glu Val Val Ala Ala Pro Gly Ser Pro Tyr His Arg Leu Leu		
	3120	3125	3130
	ACT TGT TTA CGA AAT GTC CAC AAG GTC ACC ACC TGC T GAGCGCCATG		9758
	Thr Cys Leu Arg Asn Val His Lys Val Thr Thr Cys		
	3135	3140	
	GTGGGAGAGA CTGTGAGGCG GCAGCTGGGG CCGGAGCCTT TCGAAGTCTG TGCCCTTGTG		9818
	CCCTGCCTCC ACCGAGCCAG CTTGGTCCCT ATGGGCTTCC GCACATGCCG CGGGCGGCA		9878
	CGCAACGTGC GTGTCTCTGC CATGTGGCAG AAGTGCTCTT TGTGGCAGTG GCCAGGCAGG		9938
	CAGTGTCTGC AGTCCTGGTG GGGCTGAGCC TGAGGCCTTC CAGAAAGCAG GAGCAGCTGT		9998
	CCTGCACCCC ATGTGGGTGA CCAGGTCCCT TCTCCTGATA GTCACCTGCT GGTGTGTGCC		10058
	AGGTTCCAGC TGCTCTTCCA TCTGGGCCAG AAGTCCTCCC TCCTGCAGGC TGGCTGTGG		10118
	CCCCTCTGCT GTCCTGCAGT AGAAGGTGCC GTGAGCAGCC TTTCGGGAACA CTGGCCTGGG		10178
	TCTCCCTGGT GGGGTGTCCA TGCCACGCCC CGTGTCTGGA TGCACAGATG CCATGCCCTG		10238
	TGCTGGGCA GTGGCTGGGG GTGCTAGACA CCGGCACCA TTCTCCCTTC TCTCTTTCT		10298
	TCTCAGGATT TAAATTTAA TTATATCAGT AAAGAGATTA ATTTTAACGT AAAAAAAAAA		10358
	AAAAAAAA		10366

【0112】配列番号：6

配列の長さ：3144

配列の型：アミノ酸

\*トポロジー：直鎖状

配列の種類：タンパク質

\*

Met	Ala	Thr	Leu	Glu	Lys	Leu	Met	Lys	Ala	Phe	Glu	Ser	Leu	Lys	Ser
1					5				10					15	
Phe	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln
			20					25					30		
Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro
			35					40					45		
Pro	Pro	Pro	Gln	Leu	Pro	Gln	Pro	Pro	Pro	Gln	Ala	Gln	Pro	Leu	Leu
			50				55					60			
Pro	Gln	Pro	Gln	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Gly	Pro	
			65			70				75			80		
Ala	Val	Ala	Glu	Glu	Pro	Leu	His	Arg	Pro	Lys	Lys	Glu	Leu	Ser	Ala
			85					90				95			
Thr	Lys	Lys	Asp	Arg	Val	Asn	His	Cys	Leu	Thr	Ile	Cys	Glu	Asn	Ile
			100					105				110			
Val	Ala	Gln	Ser	Val	Arg	Asn	Ser	Pro	Glu	Phe	Gln	Lys	Leu	Leu	Gly
			115					120				125			
Ile	Ala	Met	Glu	Leu	Phe	Leu	Leu	Cys	Ser	Asp	Asp	Ala	Glu	Ser	Asp



63  
 130 135 140  
 Val Arg Met Val Ala Asp Glu Cys Leu Asn Lys Val Ile Lys Ala Leu  
 145 150 155 160  
 Met Asp Ser Asn Leu Pro Arg Leu Gln Leu Glu Leu Tyr Lys Glu Ile  
 165 170 175  
 Lys Lys Asn Gly Ala Pro Arg Ser Leu Arg Ala Ala Leu Trp Arg Phe  
 180 185 190  
 Ala Glu Leu Ala His Leu Val Arg Pro Gln Lys Cys Arg Pro Tyr Leu  
 195 200 205  
 Val Asn Leu Leu Pro Cys Leu Thr Arg Thr Ser Lys Arg Pro Glu Glu  
 210 215 220  
 Ser Val Gln Glu Thr Leu Ala Ala Ala Val Pro Lys Ile Met Ala Ser  
 225 230 235 240  
 Phe Gly Asn Phe Ala Asn Asp Asn Glu Ile Lys Val Leu Leu Lys Ala  
 245 250 255  
 Phe Ile Ala Asn Leu Lys Ser Ser Ser Pro Thr Ile Arg Arg Thr Ala  
 260 265 270  
 Ala Gly Ser Ala Val Ser Ile Cys Gln His Ser Arg Arg Thr Gln Tyr  
 275 280 285  
 Phe Tyr Ser Trp Leu Leu Asn Val Leu Leu Gly Leu Leu Val Pro Val  
 290 295 300  
 Glu Asp Glu His Ser Thr Leu Leu Ile Leu Gly Val Leu Leu Thr Leu  
 305 310 315 320  
 Arg Tyr Leu Val Pro Leu Leu Gln Gln Gln Val Lys Asp Thr Ser Leu  
 325 330 335  
 Lys Gly Ser Phe Gly Val Thr Arg Lys Glu Met Glu Val Ser Pro Ser  
 340 345 350  
 Ala Glu Gln Leu Val Gln Val Tyr Glu Leu Thr Leu His His Thr Gln  
 355 360 365  
 His Gln Asp His Asn Val Val Thr Gly Ala Leu Glu Leu Leu Gln Gln  
 370 375 380  
 Leu Phe Arg Thr Pro Pro Pro Glu Leu Leu Gln Thr Leu Thr Ala Val  
 385 390 395 400  
 Gly Gly Ile Gly Gln Leu Thr Ala Ala Lys Glu Glu Ser Gly Gly Arg  
 405 410 415  
 Ser Arg Ser Gly Ser Ile Val Glu Leu Ile Ala Gly Gly Gly Ser Ser  
 420 425 430  
 Cys Ser Pro Val Leu Ser Arg Lys Gln Lys Gly Lys Val Leu Leu Gly  
 435 440 445  
 Glu Glu Glu Ala Leu Glu Asp Asp Ser Glu Ser Arg Ser Asp Val Ser  
 450 455 460  
 Ser Ser Ala Leu Thr Ala Ser Val Lys Asp Glu Ile Ser Gly Glu Leu  
 465 470 475 480  
 Ala Ala Ser Ser Gly Val Ser Thr Pro Gly Ser Ala Gly His Asp Ile  
 485 490 495  
 Ile Thr Glu Gln Pro Arg Ser Gln His Thr Leu Gln Ala Asp Ser Leu  
 500 505 510  
 Asp Leu Ala Ser Cys Asp Leu Thr Ser Ser Ala Thr Asp Gly Asp Glu  
 515 520 525  
 Glu Asp Ile Leu Ser His Ser Ser Ser Gln Val Ser Ala Val Pro Ser

65  
 530 535 540  
 Asp Pro Ala Met Asp Leu Asn Asp Gly Thr Gln Ala Ser Ser Pro Ile  
 545 550 555 560  
 Ser Asp Ser Ser Gln Thr Thr Thr Glu Gly Pro Asp Ser Ala Val Thr  
 565 570 575  
 Pro Ser Asp Ser Ser Glu Ile Val Leu Asp Gly Thr Asp Asn Gln Tyr  
 580 585 590  
 Leu Gly Leu Gln Ile Gly Gln Pro Gln Asp Glu Asp Glu Glu Ala Thr  
 595 600 605  
 Gly Ile Leu Pro Asp Glu Ala Ser Glu Ala Phe Arg Asn Ser Ser Met  
 610 615 620  
 Ala Leu Gln Gln Ala His Leu Leu Lys Asn Met Ser His Cys Arg Gln  
 625 630 635 640  
 Pro Ser Asp Ser Ser Val Asp Lys Phe Val Leu Arg Asp Glu Ala Thr  
 645 650 655  
 Glu Pro Gly Asp Gln Glu Asn Lys Pro Cys Arg Ile Lys Gly Asp Ile  
 660 665 670  
 Gly Gln Ser Thr Asp Asp Asp Ser Ala Pro Leu Val His Ser Val Arg  
 675 680 685  
 Leu Leu Ser Ala Ser Phe Leu Leu Thr Gly Gly Lys Asn Val Leu Val  
 690 695 700  
 Pro Asp Arg Asp Val Arg Val Ser Val Lys Ala Leu Ala Leu Ser Cys  
 705 710 715 720  
 Val Gly Ala Ala Val Ala Leu His Pro Glu Ser Phe Phe Ser Lys Leu  
 725 730 735  
 Tyr Lys Val Pro Leu Asp Thr Thr Glu Tyr Pro Glu Glu Gln Tyr Val  
 740 745 750  
 Ser Asp Ile Leu Asn Tyr Ile Asp His Gly Asp Pro Gln Val Arg Gly  
 755 760 765  
 Ala Thr Ala Ile Leu Cys Gly Thr Leu Ile Cys Ser Ile Leu Ser Arg  
 770 775 780  
 Ser Arg Phe His Val Gly Asp Trp Met Gly Thr Ile Arg Thr Leu Thr  
 785 790 795 800  
 Gly Asn Thr Phe Ser Leu Ala Asp Cys Ile Pro Leu Leu Arg Lys Thr  
 805 810 815  
 Leu Lys Asp Glu Ser Ser Val Thr Cys Lys Leu Ala Cys Thr Ala Val  
 820 825 830  
 Arg Asn Cys Val Met Ser Leu Cys Ser Ser Ser Tyr Ser Glu Leu Gly  
 835 840 845  
 Leu Gln Leu Ile Ile Asp Val Leu Thr Leu Arg Asn Ser Ser Tyr Trp  
 850 855 860  
 Leu Val Arg Thr Glu Leu Leu Glu Thr Leu Ala Glu Ile Asp Phe Arg  
 865 870 875 880  
 Leu Val Ser Phe Leu Glu Ala Lys Ala Glu Asn Leu His Arg Gly Ala  
 885 890 895  
 His His Tyr Thr Gly Leu Leu Lys Leu Gln Glu Arg Val Leu Asn Asn  
 900 905 910  
 Val Val Ile His Leu Leu Gly Asp Glu Asp Pro Arg Val Arg His Val  
 915 920 925  
 Ala Ala Ala Ser Leu Ile Arg Leu Val Pro Lys Leu Phe Tyr Lys Cys

67  
 930 935 940  
 Asp Gln Gly Gln Ala Asp Pro Val Val Ala Val Ala Arg Asp Gln Ser  
 945 950 955 960  
 Ser Val Tyr Leu Lys Leu Leu Met His Glu Thr Gln Pro Pro Ser His  
 965 970 975  
 Phe Ser Val Ser Thr Ile Thr Arg Ile Tyr Arg Gly Tyr Asn Leu Leu  
 980 985 990  
 Pro Ser Ile Thr Asp Val Thr Met Glu Asn Asn Leu Ser Arg Val Ile  
 995 1000 1005  
 Ala Ala Val Ser His Glu Leu Ile Thr Ser Thr Thr Arg Ala Leu Thr  
 1010 1015 1020  
 Phe Gly Cys Cys Glu Ala Leu Cys Leu Leu Ser Thr Ala Phe Pro Val  
 1025 1030 1035 1040  
 Cys Ile Trp Ser Leu Gly Trp His Cys Gly Val Pro Pro Leu Ser Ala  
 1045 1050 1055  
 Ser Asp Glu Ser Arg Lys Ser Cys Thr Val Gly Met Ala Thr Met Ile  
 1060 1065 1070  
 Leu Thr Leu Leu Ser Ser Ala Trp Phe Pro Leu Asp Leu Ser Ala His  
 1075 1080 1085  
 Gln Asp Ala Leu Ile Leu Ala Gly Asn Leu Leu Ala Ala Ser Ala Pro  
 1090 1095 1100  
 Lys Ser Leu Arg Ser Ser Trp Ala Ser Glu Glu Glu Ala Asn Pro Ala  
 1105 1110 1115 1120  
 Ala Thr Lys Gln Glu Glu Val Trp Pro Ala Leu Gly Asp Arg Ala Leu  
 1125 1130 1135  
 Val Pro Met Val Glu Gln Leu Phe Ser His Leu Leu Lys Val Ile Asn  
 1140 1145 1150  
 Ile Cys Ala His Val Leu Asp Asp Val Ala Pro Gly Pro Ala Ile Lys  
 1155 1160 1165  
 Ala Ala Leu Pro Ser Leu Thr Asn Pro Pro Ser Leu Ser Pro Ile Arg  
 1170 1175 1180  
 Arg Lys Gly Lys Glu Lys Glu Pro Gly Glu Gln Ala Ser Val Pro Leu  
 1185 1190 1195 1200  
 Ser Pro Lys Lys Gly Ser Glu Ala Ser Ala Ala Ser Arg Gln Ser Asp  
 1205 1210 1215  
 Thr Ser Gly Pro Val Thr Thr Ser Lys Ser Ser Ser Leu Gly Ser Phe  
 1220 1225 1230  
 Tyr His Leu Pro Ser Tyr Leu Arg Leu His Asp Val Leu Lys Ala Thr  
 1235 1240 1245  
 His Ala Asn Tyr Lys Val Thr Leu Asp Leu Gln Asn Ser Thr Glu Lys  
 1250 1255 1260  
 Phe Gly Gly Phe Leu Arg Ser Ala Leu Asp Val Leu Ser Gln Ile Leu  
 1265 1270 1275 1280  
 Glu Leu Ala Thr Leu Gln Asp Ile Gly Lys Cys Val Glu Glu Ile Leu  
 1285 1290 1295  
 Gly Tyr Leu Lys Ser Cys Phe Ser Arg Glu Pro Met Met Ala Thr Val  
 1300 1305 1310  
 Cys Val Gln Gln Leu Leu Lys Thr Leu Phe Gly Thr Asn Leu Ala Ser  
 1315 1320 1325  
 Gln Phe Asp Gly Leu Ser Ser Asn Pro Ser Lys Ser Gln Gly Arg Ala



1330	1335	1340	
Gln Arg Leu Gly Ser Ser Ser Val Arg Pro Gly Leu Tyr His Tyr Cys			
1345	1350	1355	1360
Phe Met Ala Pro Tyr Thr His Phe Thr Gln Ala Leu Ala Asp Ala Ser			
	1365	1370	1375
Leu Arg Asn Met Val Gln Ala Glu Gln Glu Asn Asp Thr Ser Gly Trp			
	1380	1385	1390
Phe Asp Val Leu Gln Lys Val Ser Thr Gln Leu Lys Thr Asn Leu Thr			
	1395	1400	1405
Ser Val Thr Lys Asn Arg Ala Asp Lys Asn Ala Ile His Asn His Ile			
	1410	1415	1420
Arg Leu Phe Glu Pro Leu Val Ile Lys Ala Leu Lys Gln Tyr Thr Thr			
1425	1430	1435	1440
Thr Thr Cys Val Gln Leu Gln Lys Gln Val Leu Asp Leu Leu Ala Gln			
	1445	1450	1455
Leu Val Gln Leu Arg Val Asn Tyr Cys Leu Leu Asp Ser Asp Gln Val			
	1460	1465	1470
Phe Ile Gly Phe Val Leu Lys Gln Phe Glu Tyr Ile Glu Val Gly Gln			
	1475	1480	1485
Phe Arg Glu Ser Glu Ala Ile Ile Pro Asn Ile Phe Phe Phe Leu Val			
	1490	1495	1500
Leu Leu Ser Tyr Glu Arg Tyr His Ser Lys Gln Ile Ile Gly Ile Pro			
1505	1510	1515	1520
Lys Ile Ile Gln Leu Cys Asp Gly Ile Met Ala Ser Gly Arg Lys Ala			
	1525	1530	1535
Val Thr His Ala Ile Pro Ala Leu Gln Pro Ile Val His Asp Leu Phe			
	1540	1545	1550
Val Leu Arg Gly Thr Asn Lys Ala Asp Ala Gly Lys Glu Leu Glu Thr			
	1555	1560	1565
Gln Lys Glu Val Val Val Ser Met Leu Leu Arg Leu Ile Gln Tyr His			
	1570	1575	1580
Gln Val Leu Glu Met Phe Ile Leu Val Leu Gln Gln Cys His Lys Glu			
1585	1590	1595	1600
Asn Glu Asp Lys Trp Lys Arg Leu Ser Arg Gln Ile Ala Asp Ile Ile			
	1605	1610	1615
Leu Pro Met Leu Ala Lys Gln Gln Met His Ile Asp Ser His Glu Ala			
	1620	1625	1630
Leu Gly Val Leu Asn Thr Leu Phe Glu Ile Leu Ala Pro Ser Ser Leu			
	1635	1640	1645
Arg Pro Val Asp Met Leu Leu Arg Ser Met Phe Val Thr Pro Asn Thr			
	1650	1655	1660
Met Ala Ser Val Ser Thr Val Gln Leu Trp Ile Ser Gly Ile Leu Ala			
1665	1670	1675	1680
Ile Leu Arg Val Leu Ile Ser Gln Ser Thr Glu Asp Ile Val Leu Ser			
	1685	1690	1695
Arg Ile Gln Glu Leu Ser Phe Ser Pro Tyr Leu Ile Ser Cys Thr Val			
	1700	1705	1710
Ile Asn Arg Leu Arg Asp Gly Asp Ser Thr Ser Thr Leu Glu Glu His			
	1715	1720	1725
Ser Glu Gly Lys Gln Ile Lys Asn Leu Pro Glu Glu Thr Phe Ser Arg			

71

72

1730	1735	1740	
Phe Leu Leu Gln Leu Val Gly Ile Leu Leu Glu Asp Ile Val Thr Lys			
1745	1750	1755	1760
Gln Leu Lys Val Glu Met Ser Glu Gln Gln His Thr Phe Tyr Cys Gln			
	1765	1770	1775
Glu Leu Gly Thr Leu Leu Met Cys Leu Ile His Ile Phe Lys Ser Gly			
	1780	1785	1790
Met Phe Arg Arg Ile Thr Ala Ala Ala Thr Arg Leu Phe Arg Ser Asp			
	1795	1800	1805
Gly Cys Gly Gly Ser Phe Tyr Thr Leu Asp Ser Leu Asn Leu Arg Ala			
	1810	1815	1820
Arg Ser Met Ile Thr Thr His Pro Ala Leu Val Leu Leu Trp Cys Gln			
1825	1830	1835	1840
Ile Leu Leu Leu Val Asn His Thr Asp Tyr Arg Trp Trp Ala Glu Val			
	1845	1850	1855
Gln Gln Thr Pro Lys Arg His Ser Leu Ser Ser Thr Lys Leu Leu Ser			
	1860	1865	1870
Pro Gln Met Ser Gly Glu Glu Glu Asp Ser Asp Leu Ala Ala Lys Leu			
	1875	1880	1885
Gly Met Cys Asn Arg Glu Ile Val Arg Arg Gly Ala Leu Ile Leu Phe			
	1890	1895	1900
Cys Asp Tyr Val Cys Gln Asn Leu His Asp Ser Glu His Leu Thr Trp			
1905	1910	1915	1920
Leu Ile Val Asn His Ile Gln Asp Leu Ile Ser Leu Ser His Glu Pro			
	1925	1930	1935
Pro Val Gln Asp Phe Ile Ser Ala Val His Arg Asn Ser Ala Ala Ser			
	1940	1945	1950
Gly Leu Phe Ile Gln Ala Ile Gln Ser Arg Cys Glu Asn Leu Ser Thr			
	1955	1960	1965
Pro Thr Met Leu Lys Lys Thr Leu Gln Cys Leu Glu Gly Ile His Leu			
	1970	1975	1980
Ser Gln Ser Gly Ala Val Leu Thr Leu Tyr Val Asp Arg Leu Leu Cys			
1985	1990	1995	2000
Thr Pro Phe Arg Val Leu Ala Arg Met Val Asp Ile Leu Ala Cys Arg			
	2005	2010	2015
Arg Val Glu Met Leu Leu Ala Ala Asn Leu Gln Ser Ser Met Ala Gln			
	2020	2025	2030
Leu Pro Met Glu Glu Leu Asn Arg Ile Gln Glu Tyr Leu Gln Ser Ser			
	2035	2040	2045
Gly Leu Ala Gln Arg His Gln Arg Leu Tyr Ser Leu Leu Asp Arg Phe			
	2050	2055	2060
Arg Leu Ser Thr Met Gln Asp Ser Leu Ser Pro Ser Pro Pro Val Ser			
2065	2070	2075	2080
Ser His Pro Leu Asp Gly Asp Gly His Val Ser Leu Glu Thr Val Ser			
	2085	2090	2095
Pro Asp Lys Asp Trp Tyr Val His Leu Val Lys Ser Gln Cys Trp Thr			
	2100	2105	2110
Arg Ser Asp Ser Ala Leu Leu Glu Gly Ala Glu Leu Val Asn Arg Ile			
	2115	2120	2125
Pro Ala Glu Asp Met Asn Ala Phe Met Met Asn Ser Glu Phe Asn Leu			

73  
 2130 2135 2140  
 Ser Leu Leu Ala Pro Cys Leu Ser Leu Gly Met Ser Glu Ile Ser Gly  
 2145 2150 2155 2160  
 Gly Gln Lys Ser Ala Leu Phe Glu Ala Ala Arg Glu Val Thr Leu Ala  
 2165 2170 2175  
 Arg Val Ser Gly Thr Val Gln Gln Leu Pro Ala Val His His Val Phe  
 2180 2185 2190  
 Gln Pro Glu Leu Pro Ala Glu Pro Ala Ala Tyr Trp Ser Lys Leu Asn  
 2195 2200 2205  
 Asp Leu Phe Gly Asp Ala Ala Leu Tyr Gln Ser Leu Pro Thr Leu Ala  
 2210 2215 2220  
 Arg Ala Leu Ala Gln Tyr Leu Val Val Val Ser Lys Leu Pro Ser His  
 2225 2230 2235 2240  
 Leu His Leu Pro Pro Glu Lys Glu Lys Asp Ile Val Lys Phe Val Val  
 2245 2250 2255  
 Ala Thr Leu Glu Ala Leu Ser Trp His Leu Ile His Glu Gln Ile Pro  
 2260 2265 2270  
 Leu Ser Leu Asp Leu Gln Ala Gly Leu Asp Cys Cys Cys Leu Ala Leu  
 2275 2280 2285  
 Gln Leu Pro Gly Leu Trp Ser Val Val Ser Ser Thr Glu Phe Val Thr  
 2290 2295 2300  
 His Ala Cys Ser Leu Ile Tyr Cys Val His Phe Ile Leu Glu Ala Val  
 2305 2310 2315 2320  
 Ala Val Gln Pro Gly Glu Gln Leu Leu Ser Pro Glu Arg Arg Thr Asn  
 2325 2330 2335  
 Thr Pro Lys Ala Ile Ser Glu Glu Glu Glu Glu Val Asp Pro Asn Thr  
 2340 2345 2350  
 Gln Asn Pro Lys Tyr Ile Thr Ala Ala Cys Glu Met Val Ala Glu Met  
 2355 2360 2365  
 Val Glu Ser Leu Gln Ser Val Leu Ala Leu Gly His Lys Arg Asn Ser  
 2370 2375 2380  
 Gly Val Pro Ala Phe Leu Thr Pro Leu Leu Arg Asn Ile Ile Ile Ser  
 2385 2390 2395 2400  
 Leu Ala Arg Leu Pro Leu Val Asn Ser Tyr Thr Arg Val Pro Pro Leu  
 2405 2410 2415  
 Val Trp Lys Leu Gly Trp Ser Pro Lys Pro Gly Gly Asp Phe Gly Thr  
 2420 2425 2430  
 Ala Phe Pro Glu Ile Pro Val Glu Phe Leu Gln Glu Lys Glu Val Phe  
 2435 2440 2445  
 Lys Glu Phe Ile Tyr Arg Ile Asn Thr Leu Gly Trp Thr Ser Arg Thr  
 2450 2455 2460  
 Gln Phe Glu Glu Thr Trp Ala Thr Leu Leu Gly Val Leu Val Thr Gln  
 2465 2470 2475 2480  
 Pro Leu Val Met Glu Gln Glu Glu Ser Pro Pro Glu Glu Asp Thr Glu  
 2485 2490 2495  
 Arg Thr Gln Ile Asn Val Leu Ala Val Gln Ala Ile Thr Ser Leu Val  
 2500 2505 2510  
 Leu Ser Ala Met Thr Val Pro Val Ala Gly Asn Pro Ala Val Ser Cys  
 2515 2520 2525  
 Leu Glu Gln Gln Pro Arg Asn Lys Pro Leu Lys Ala Leu Asp Thr Arg



75

76

2530	2535	2540	
Phe Gly Arg Lys Leu Ser Ile Ile Arg Gly Ile Val Glu Gln Glu Ile			
2545	2550	2555	2560
Gln Ala Met Val Ser Lys Arg Glu Asn Ile Ala Thr His His Leu Tyr			
	2565	2570	2575
Gln Ala Trp Asp Pro Val Pro Ser Leu Ser Pro Ala Thr Thr Gly Ala			
	2580	2585	2590
Leu Ile Ser His Glu Lys Leu Leu Leu Gln Ile Asn Pro Glu Arg Glu			
	2595	2600	2605
Leu Gly Ser Met Ser Tyr Lys Leu Gly Gln Val Ser Ile His Ser Val			
	2610	2615	2620
Trp Leu Gly Asn Ser Ile Thr Pro Leu Arg Glu Glu Glu Trp Asp Glu			
2625	2630	2635	2640
Glu Glu Glu Glu Glu Ala Asp Ala Pro Ala Pro Ser Ser Pro Pro Thr			
	2645	2650	2655
Ser Pro Val Asn Ser Arg Lys His Arg Ala Gly Val Asp Ile His Ser			
	2660	2665	2670
Cys Ser Gln Phe Leu Leu Glu Leu Tyr Ser Arg Trp Ile Leu Pro Ser			
	2675	2680	2685
Ser Ser Ala Arg Arg Thr Pro Ala Ile Leu Ile Ser Glu Val Val Arg			
	2690	2695	2700
Ser Leu Leu Val Val Ser Asp Leu Phe Thr Glu Arg Asn Gln Phe Glu			
2705	2710	2715	2720
Leu Met Tyr Val Thr Leu Thr Glu Leu Arg Arg Val His Pro Ser Glu			
	2725	2730	2735
Asp Glu Ile Leu Ala Gln Tyr Leu Val Pro Ala Thr Cys Lys Ala Ala			
	2740	2745	2750
Ala Val Leu Gly Met Asp Lys Ala Val Ala Glu Pro Val Ser Arg Leu			
	2755	2760	2765
Leu Glu Ser Thr Leu Arg Ser Ser His Leu Pro Ser Arg Val Gly Ala			
	2770	2775	2780
Leu His Gly Ile Leu Tyr Val Leu Glu Cys Asp Leu Leu Asp Asp Thr			
2785	2790	2795	2800
Ala Lys Gln Leu Ile Pro Val Ile Ser Asp Tyr Leu Leu Ser Asn Leu			
	2805	2810	2815
Lys Gly Ile Ala His Cys Val Asn Ile His Ser Gln Gln His Val Leu			
	2820	2825	2830
Val Met Cys Ala Thr Ala Phe Tyr Leu Ile Glu Asn Tyr Pro Leu Asp			
	2835	2840	2845
Val Gly Pro Glu Phe Ser Ala Ser Ile Ile Gln Met Cys Gly Val Met			
	2850	2855	2860
Leu Ser Gly Ser Glu Glu Ser Thr Pro Ser Ile Ile Tyr His Cys Ala			
2865	2870	2875	2880
Leu Arg Gly Leu Glu Arg Leu Leu Leu Ser Glu Gln Leu Ser Arg Leu			
	2885	2890	2895
Asp Ala Glu Ser Leu Val Lys Leu Ser Val Asp Arg Val Asn Val His			
	2900	2905	2910
Ser Pro His Arg Ala Met Ala Ala Leu Gly Leu Met Leu Thr Cys Met			
	2915	2920	2925
Tyr Thr Gly Lys Glu Lys Val Ser Pro Gly Arg Thr Ser Asp Pro Asn			

77  
2930 2935 2940  
Pro Ala Ala Pro Asp Ser Glu Ser Val Ile Val Ala Met Glu Arg Val  
2945 2950 2955 2960  
Ser Val Leu Phe Asp Arg Ile Arg Lys Gly Phe Pro Cys Glu Ala Arg  
2965 2970 2975  
Val Val Ala Arg Ile Leu Pro Gln Phe Leu Asp Asp Phe Phe Pro Pro  
2980 2985 2990  
Gln Asp Ile Met Asn Lys Val Ile Gly Glu Phe Leu Ser Asn Gln Gln  
2995 3000 3005  
Pro Tyr Pro Gln Phe Met Ala Thr Val Val Tyr Lys Val Phe Gln Thr  
3010 3015 3020  
Leu His Ser Thr Gly Gln Ser Ser Met Val Arg Asp Trp Val Met Leu  
3025 3030 3035 3040  
Ser Leu Ser Asn Phe Thr Gln Arg Ala Pro Val Ala Met Ala Thr Trp  
3045 3050 3055  
Ser Leu Ser Cys Phe Phe Val Ser Ala Ser Thr Ser Pro Trp Val Ala  
3060 3065 3070  
Ala Ile Leu Pro His Val Ile Ser Arg Met Gly Lys Leu Glu Gln Val  
3075 3080 3085  
Asp Val Asn Leu Phe Cys Leu Val Ala Thr Asp Phe Tyr Arg His Gln  
3090 3095 3100  
Ile Glu Glu Glu Leu Asp Arg Arg Ala Phe Gln Ser Val Leu Glu Val  
3105 3110 3115 3120  
Val Ala Ala Pro Gly Ser Pro Tyr His Arg Leu Leu Thr Cys Leu Arg  
3125 3130 3135  
Asn Val His Lys Val Thr Thr Cys  
3140

## 【図面の簡単な説明】

【図1】 HD候補領域の長い範囲の制限地図である。

【図2】 ハンチンチン（IT15）転写体のノーザン  
ブロット分析の結果を示す電気泳動の図面に代わる写真  
である。【図3】 IT15転写体を規定するcDNAクローンの  
模式図である。【図4】 ハンチンチン（IT15）の混成配列の前段  
部分である。【図5】 ハンチンチン（IT15）の混成配列の中段  
部分である。【図6】 ハンチンチン（IT15）の混成配列の後段  
部分である。【図7】 (CAG)<sub>n</sub> 反復のDNA配列分析の結果を示  
す電気泳動の図面に代わる写真である。【図8】 若年期症状を示す幾つかの子供を用いたベネ  
ズエラHD同胞群の(CAG)<sub>n</sub> 反復のPCR分析の結果  
を示す電気泳動の図面に代わる写真である。【図9】 同じHD単相型についてホモ接合の子孫を用  
いたベネズエラHD同胞群の(CAG)<sub>n</sub> 反復のPCR分  
析の結果を示す電気泳動の図面に代わる写真である。【図10】 主要なHD単相型についてホモ接合の個体  
を用いたアメリカ族の一員における(CAG)<sub>n</sub> 反復のPCR分析の結果を示す電気泳動の図面に代わる写真であ  
る。30 【図11】 HDを引き起こすと考えられる新たな突然  
変異による2つのファミリーの(CAG)<sub>n</sub> 反復のPCR  
分析の結果を示す電気泳動の図面に代わる写真である。【図12】 HDを引き起こすと考えられる新たな突然  
変異による2つのファミリーの(CAG)<sub>n</sub> 反復のPCR  
分析の結果を示す電気泳動の図面に代わる写真である。【図13】 対照染色体及びHD染色体上の(CAG)<sub>n</sub>  
反復単位の数の比較を示すグラフである。【図14】 母親及び父親から伝達されるHD染色体上  
の(CAG)<sub>n</sub> 反復単位の数の比較を示すグラフである。40 【図15】 母親及び父親から伝達されるHD染色体上  
の(CAG)<sub>n</sub> 反復単位の数の比較を示すグラフである。【図16】 3つの巨大なファミリー由来のHD染色体  
上の(CAG)<sub>n</sub> 反復単位の数を別のHD発端者と比較し  
たグラフである。【図17】 3つの巨大なファミリー由来のHD染色体  
上の(CAG)<sub>n</sub> 反復単位の数を別のHD発端者と比較し  
たグラフである。【図18】 3つの巨大なファミリー由来のHD染色体  
上の(CAG)<sub>n</sub> 反復単位の数を別のHD発端者と比較し  
たグラフである。

50

【図19】 両親及び対応する子供の(CAG)<sub>n</sub> 反復長の関係を示すグラフである。

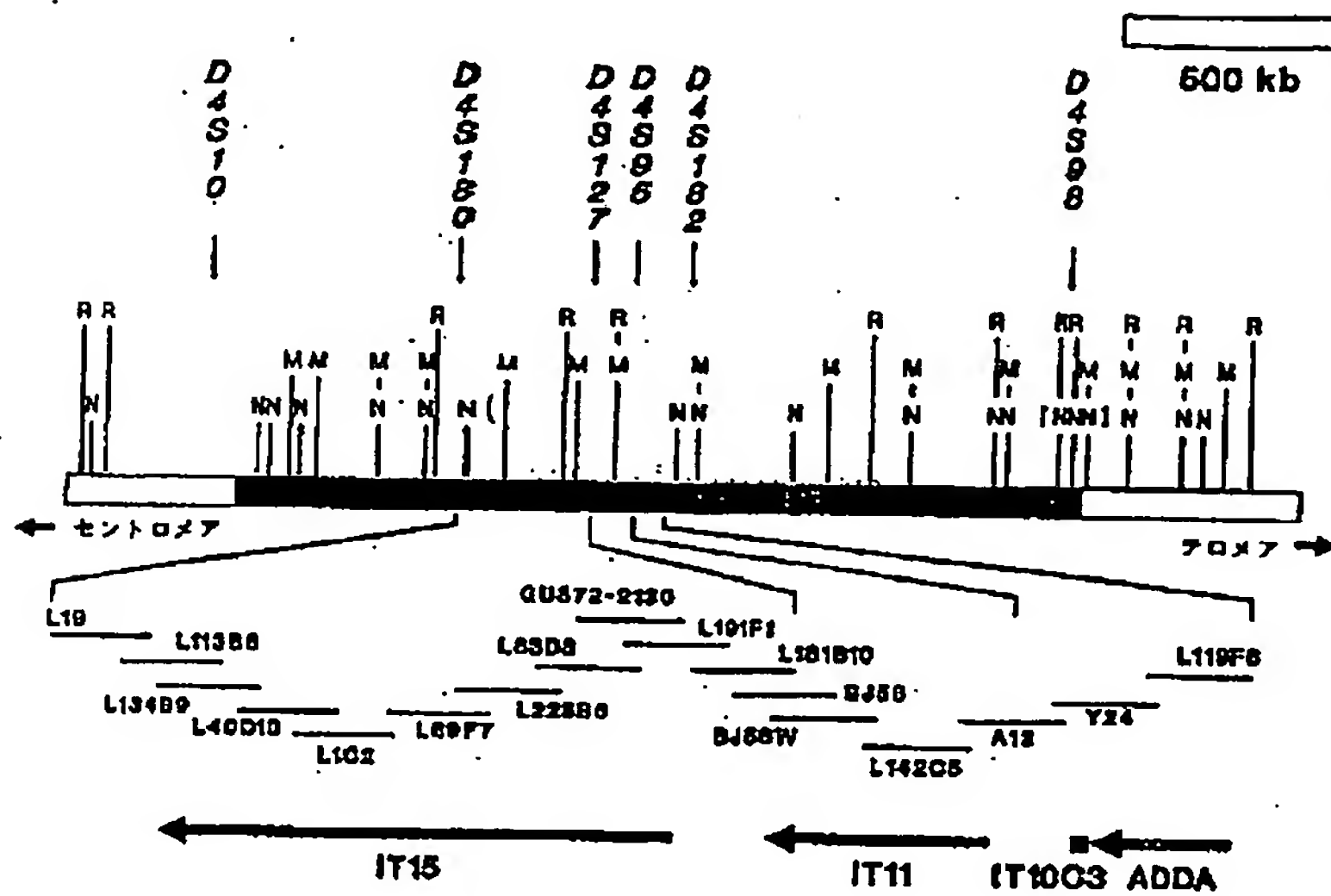
【図20】 両親及び対応する子供の(CAG)<sub>n</sub> 反復長の関係を示すグラフである。

【図21】 精子及びリンパ芽球DNA由来のHD (C\*

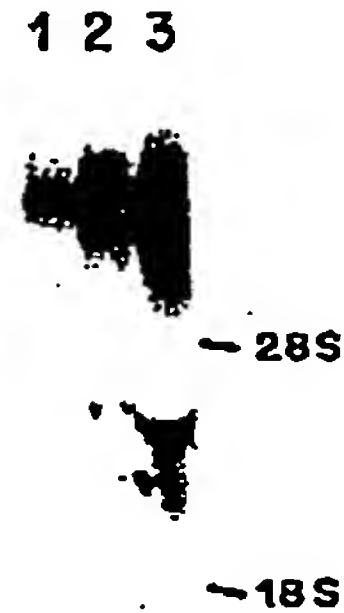
\* AG)<sub>n</sub> 反復の増幅を示す電気泳動の図面に代わる写真である。

【図22】 発症年齢と反復単位の長さとの関係を示すグラフである。

【図1】



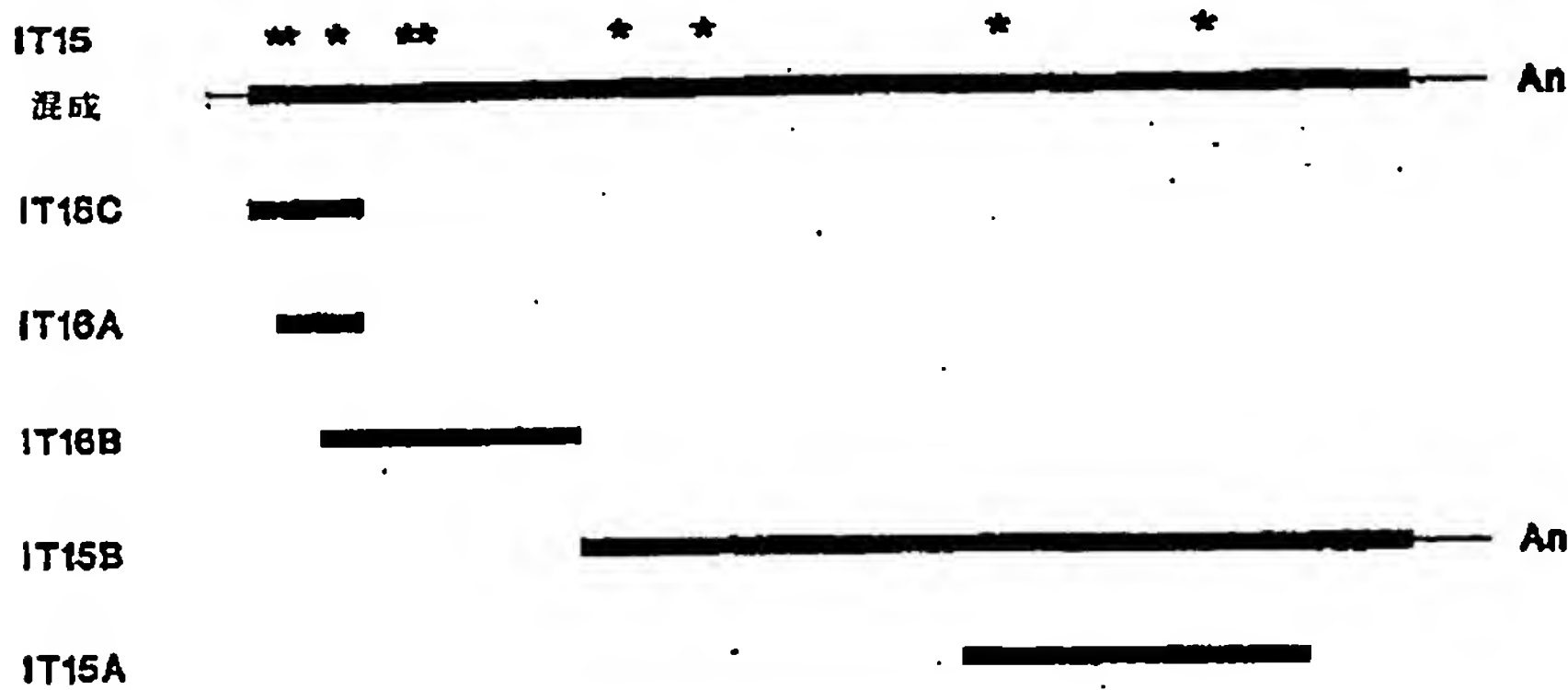
【図2】



【図7】



【図3】





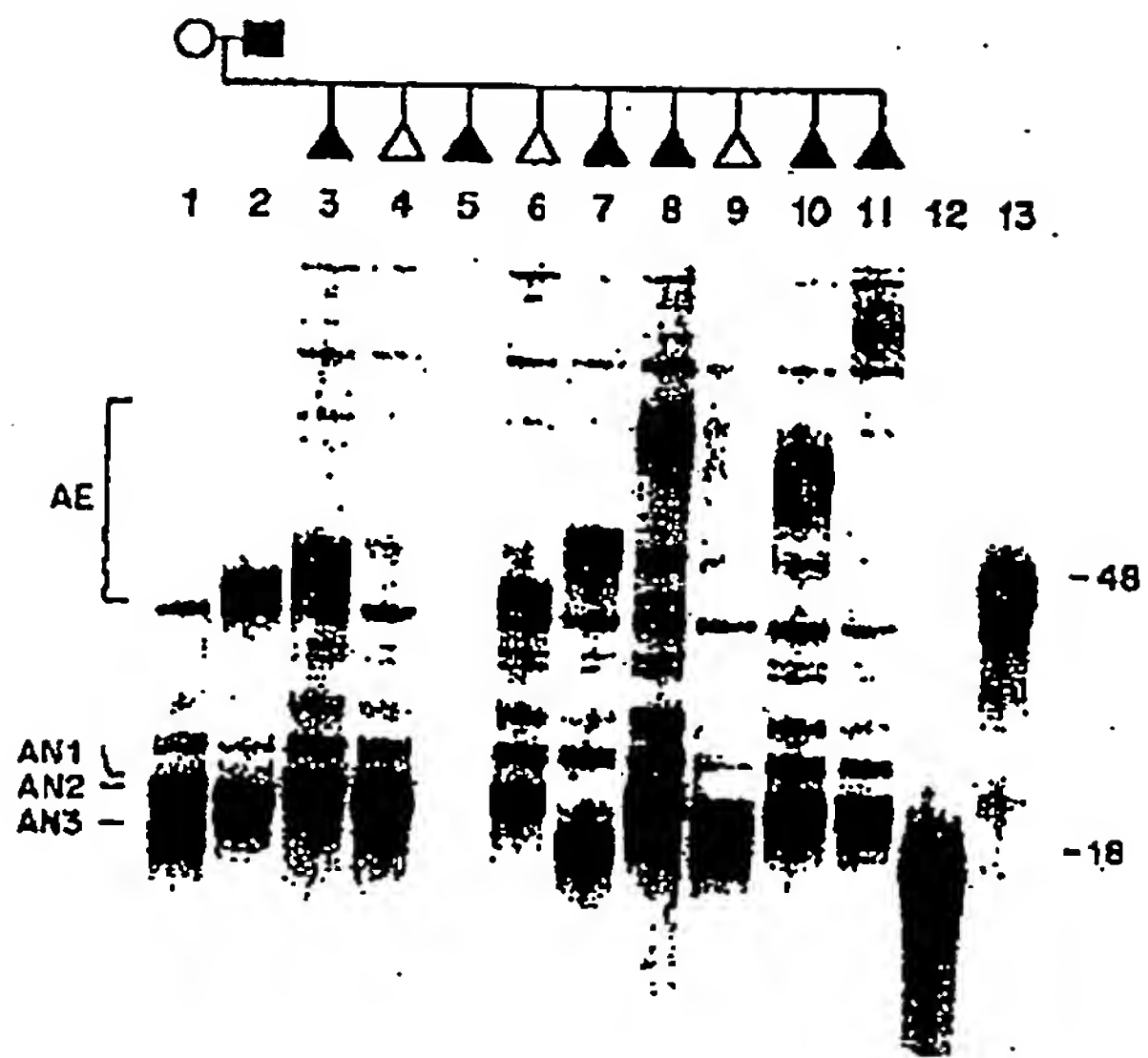
[illegible]

[illegible]

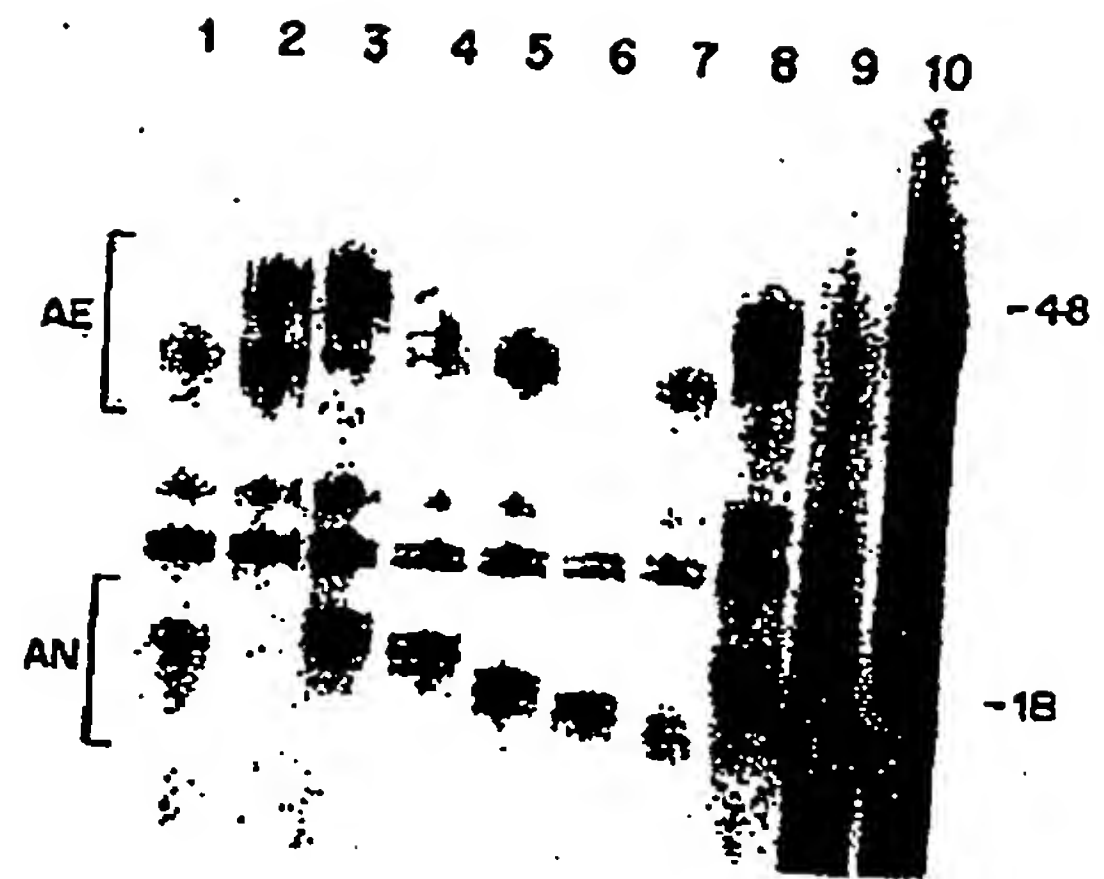
[illegible]



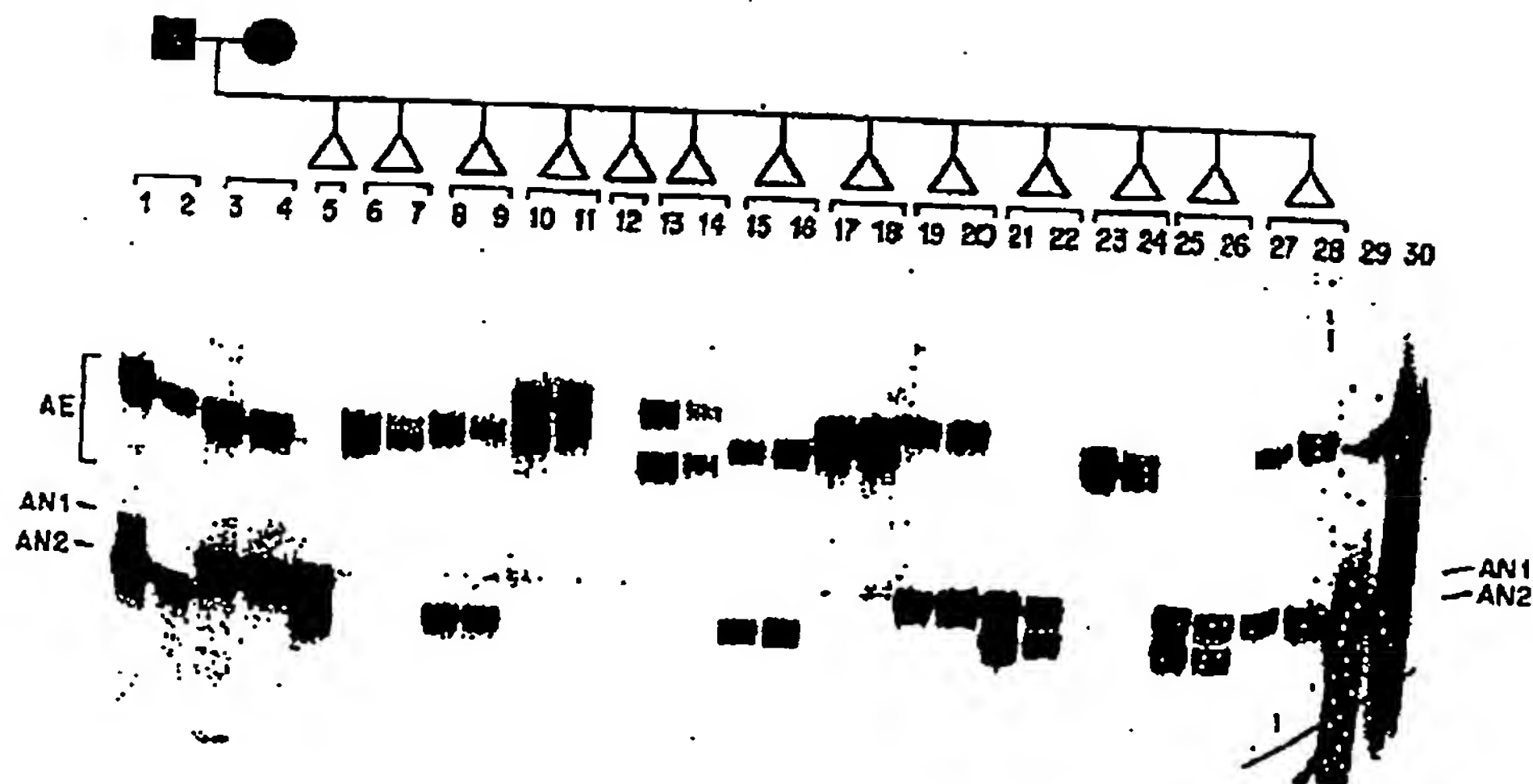
【図8】



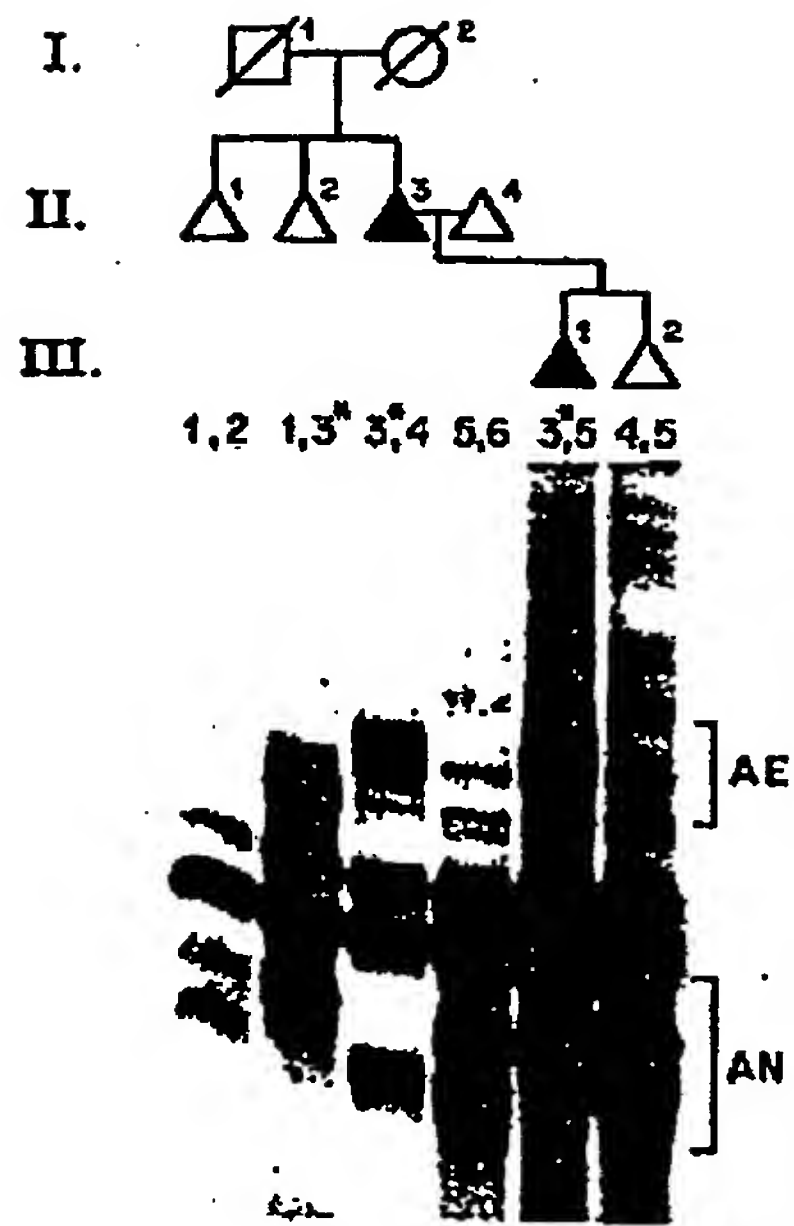
【図10】



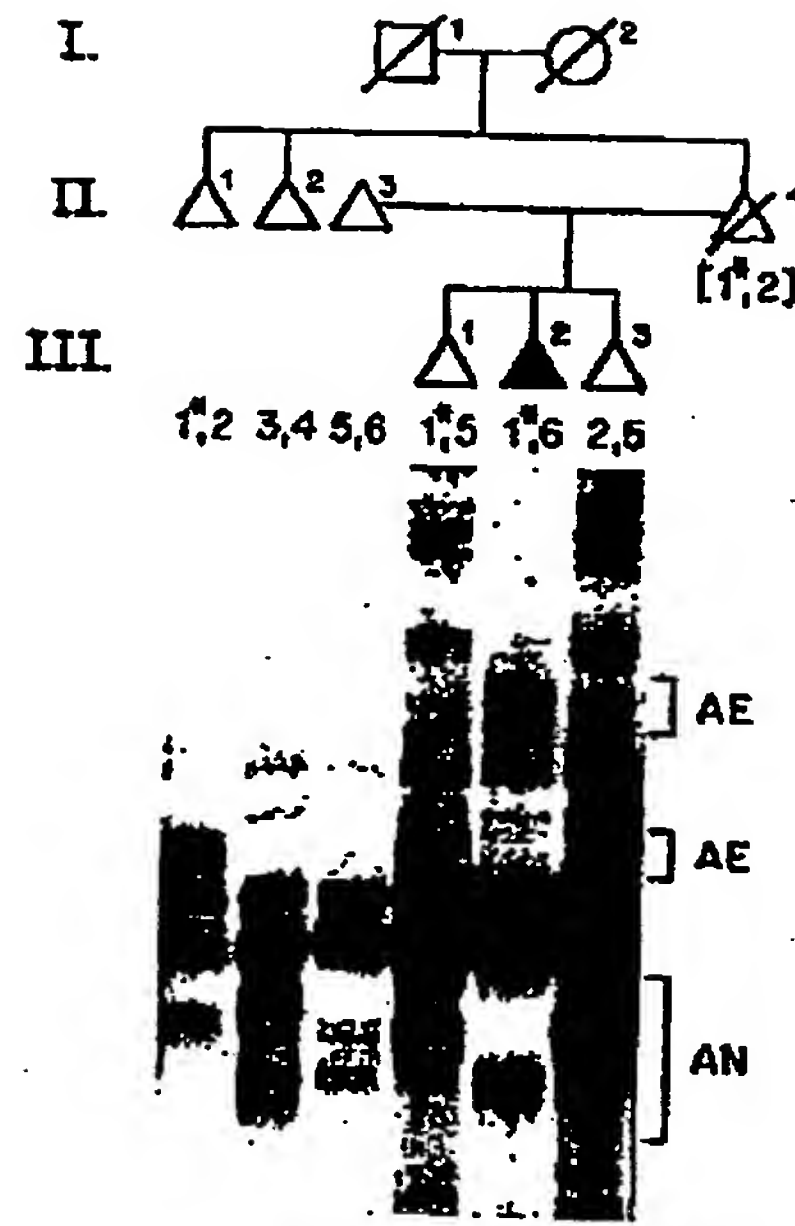
【図9】



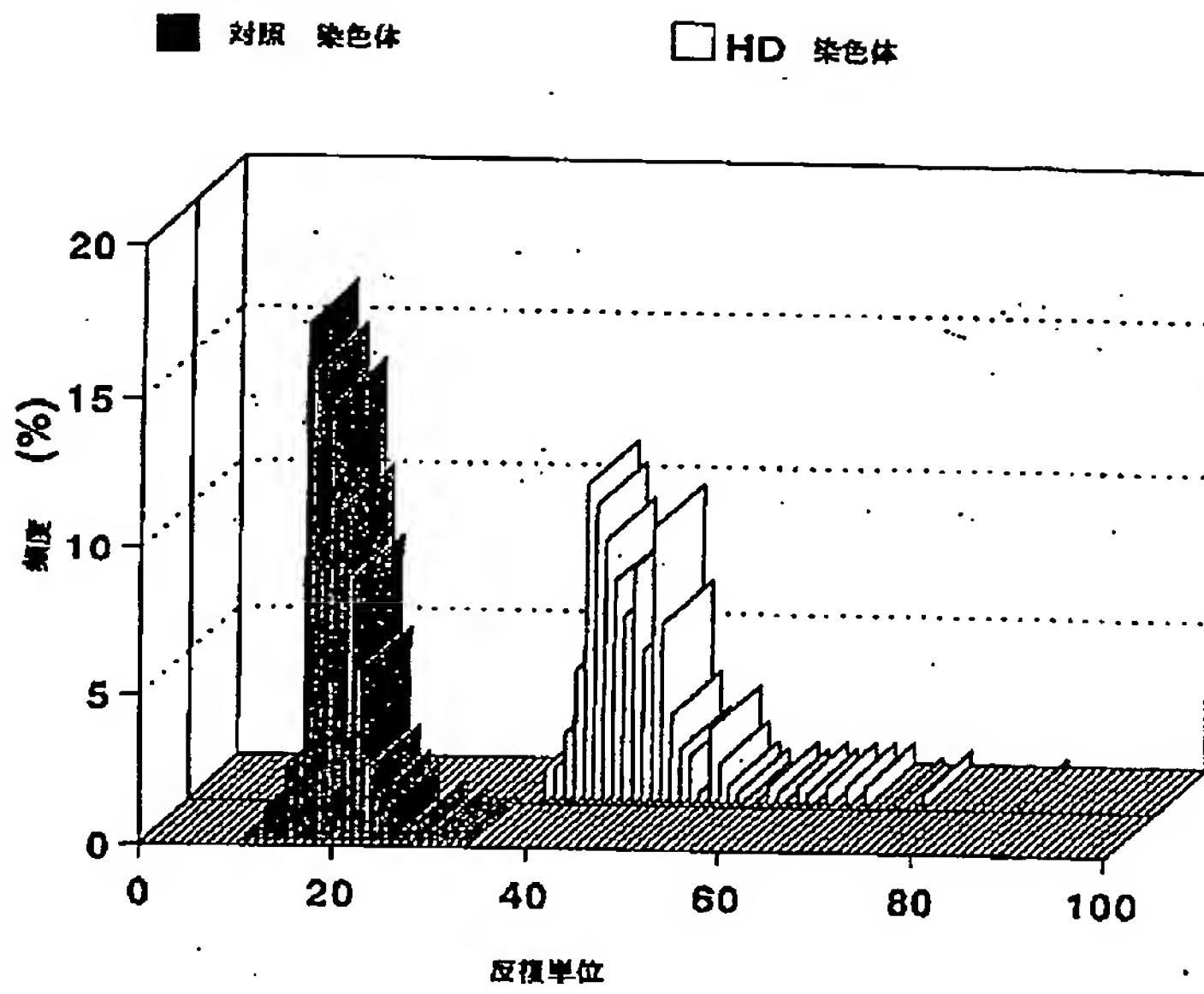
【図11】



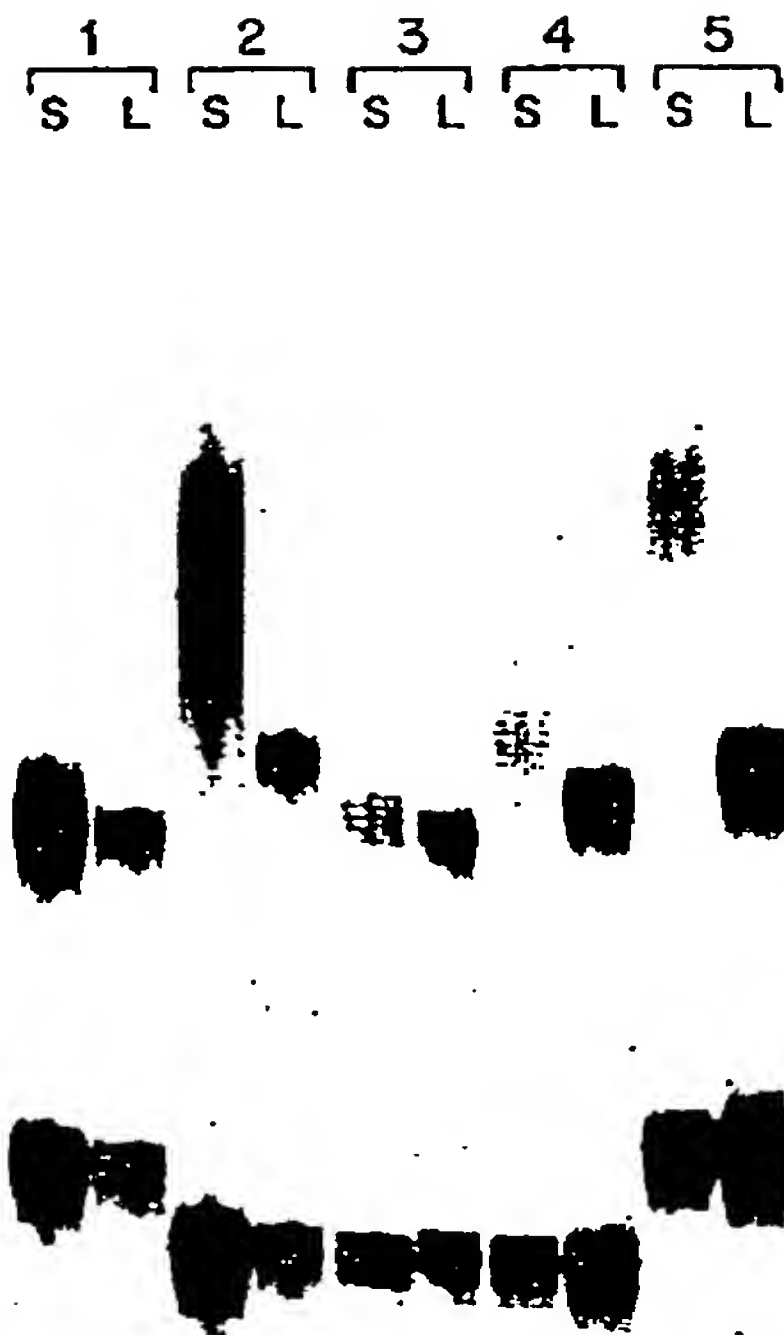
【図12】



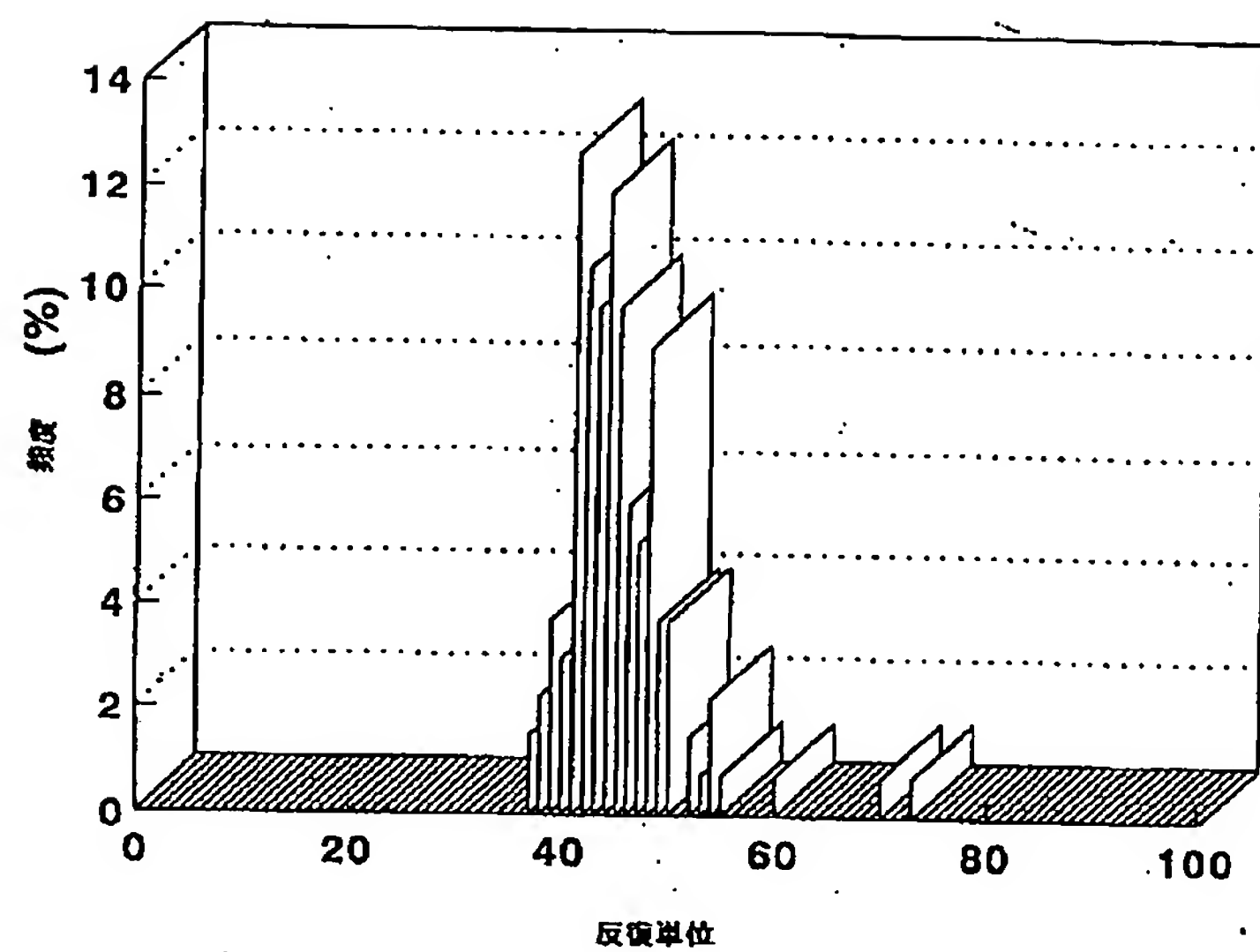
【図13】



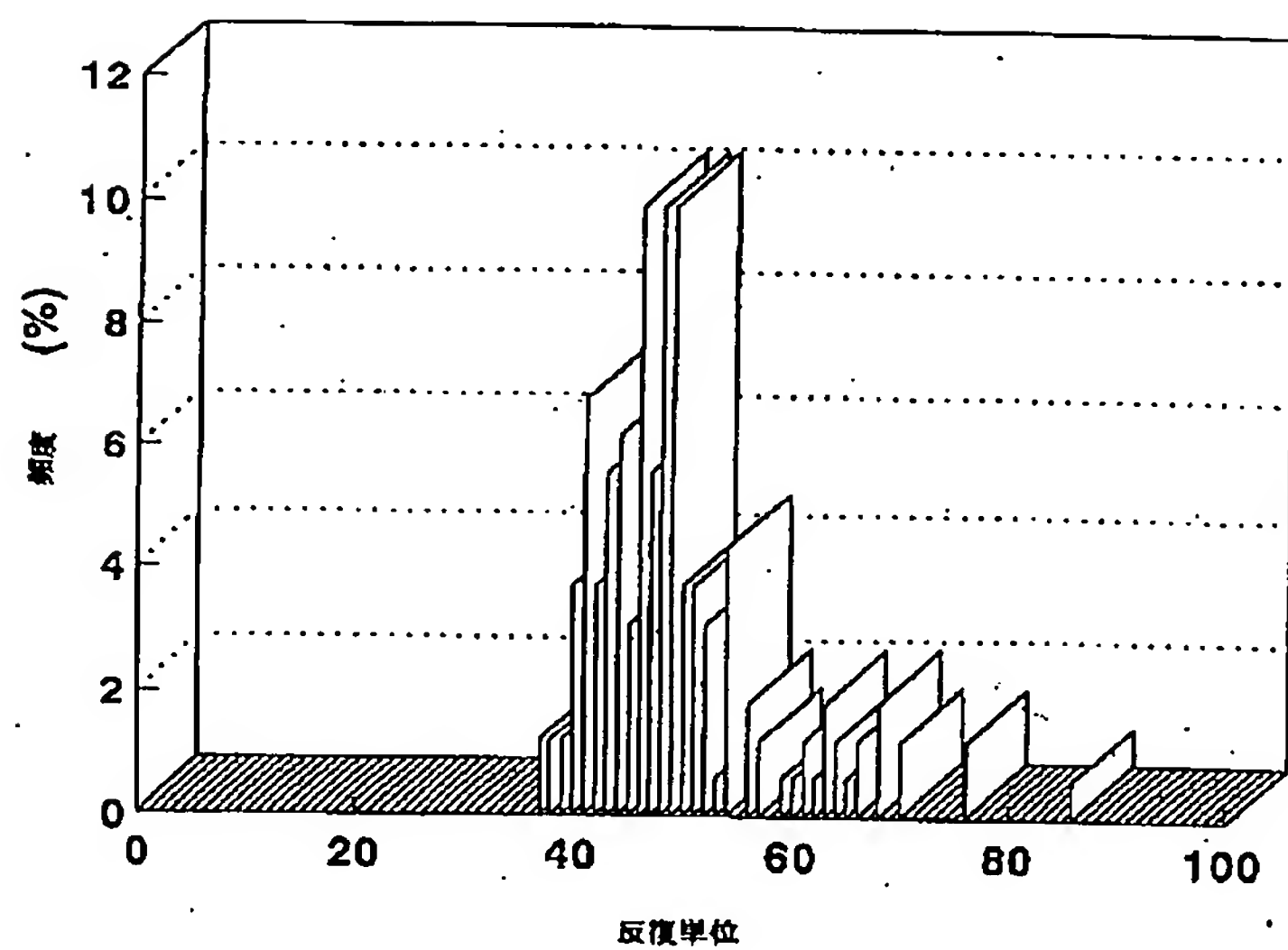
【図21】



【図14】

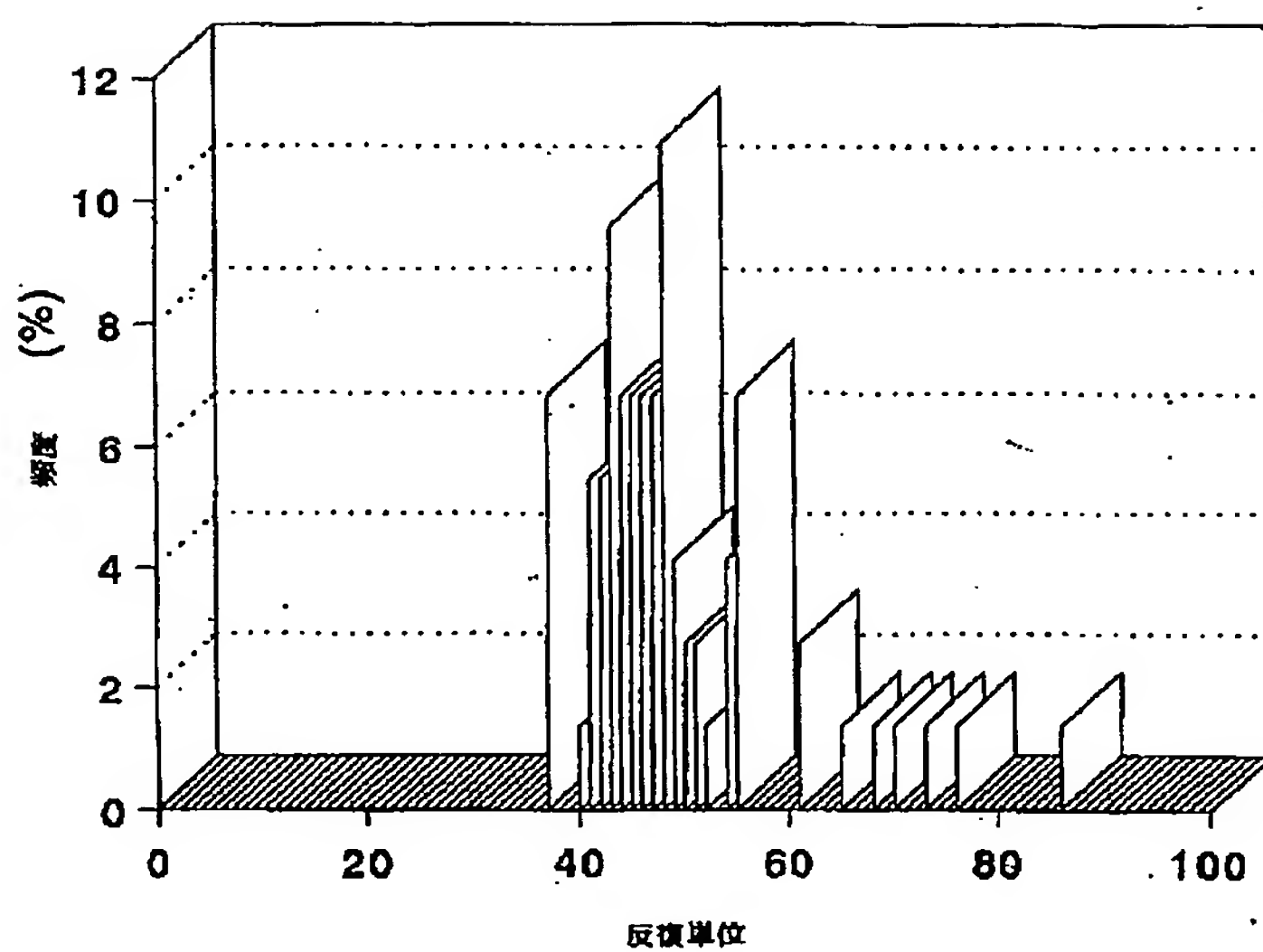


【図15】

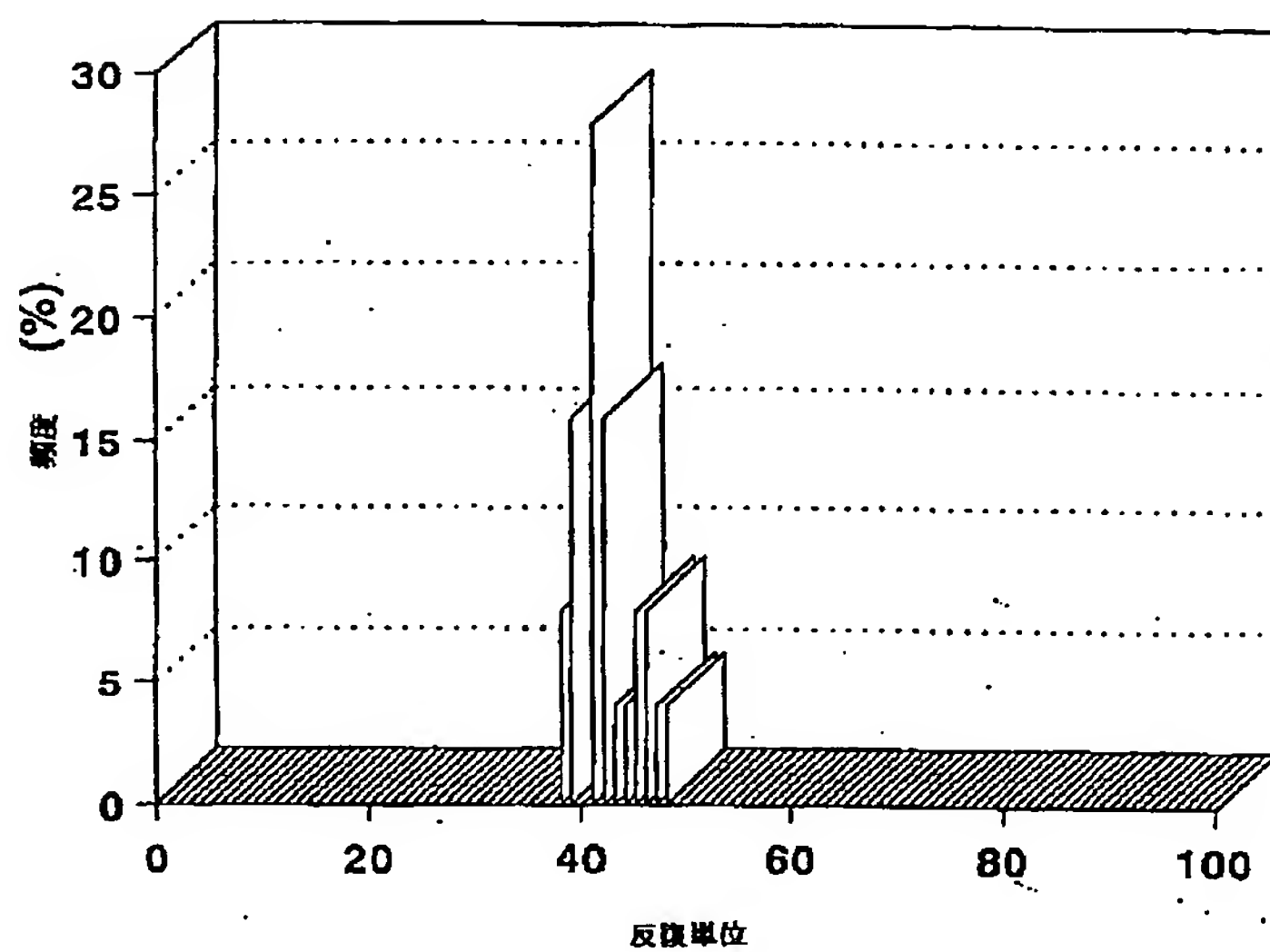




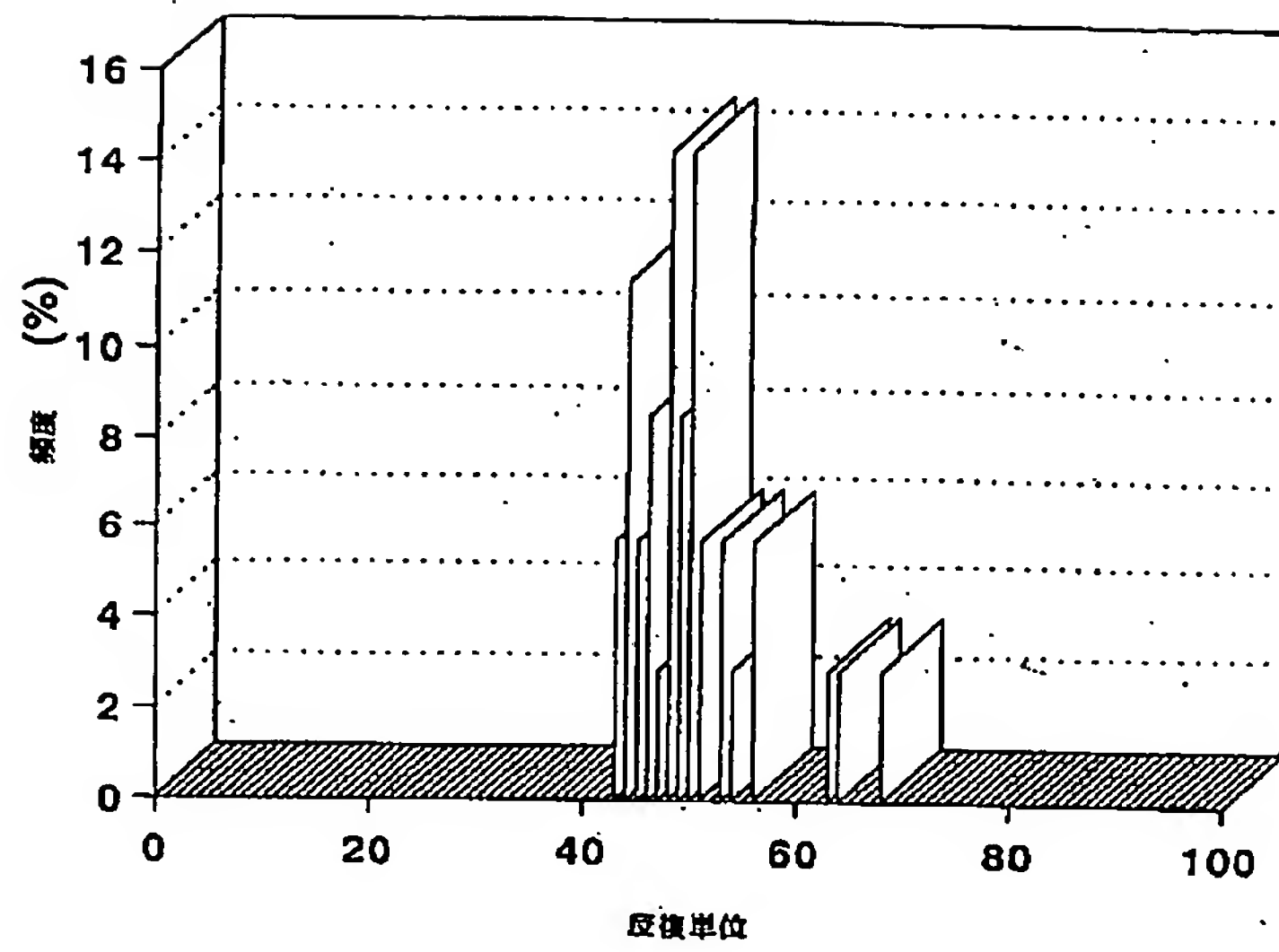
【図16】



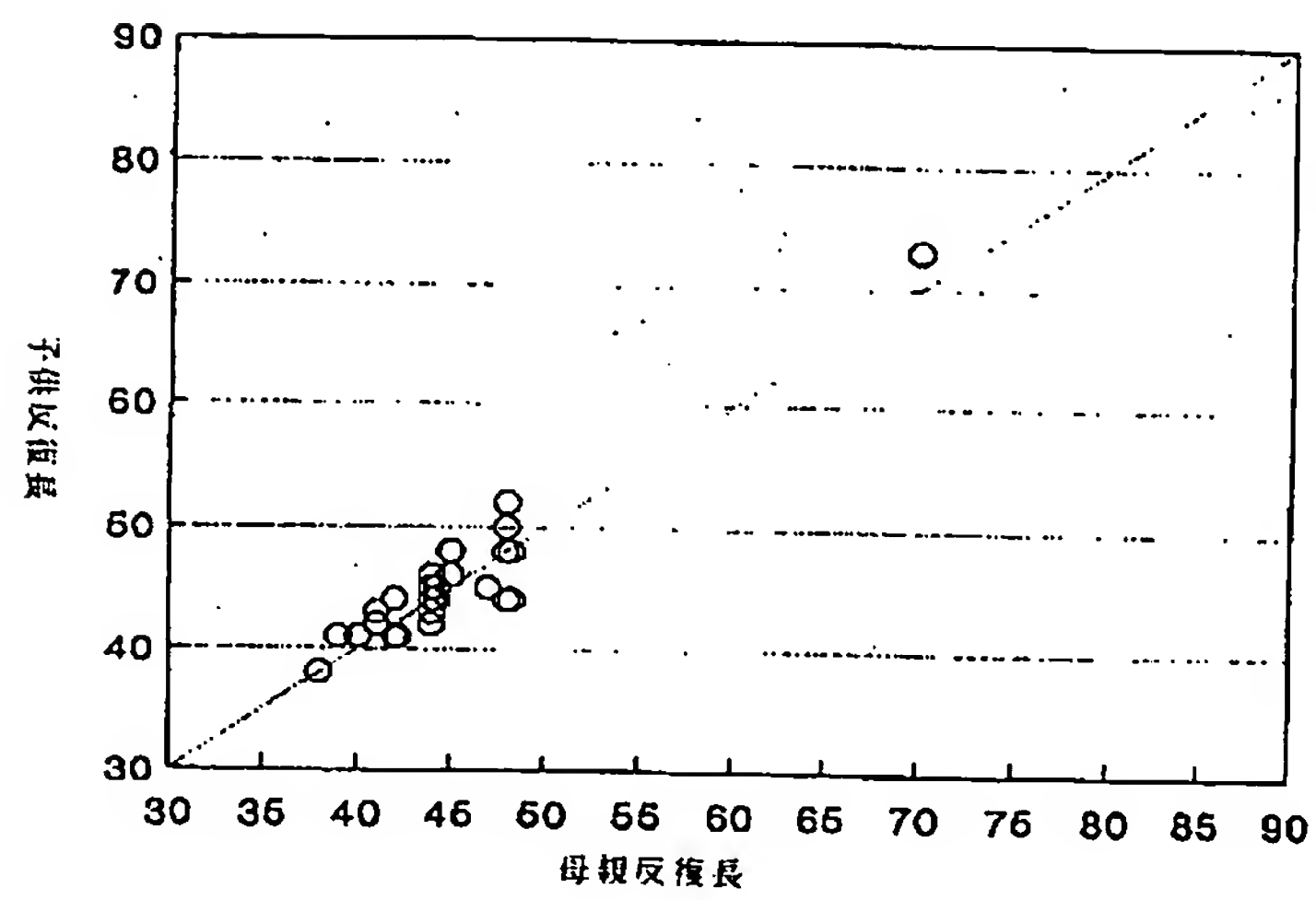
【図17】



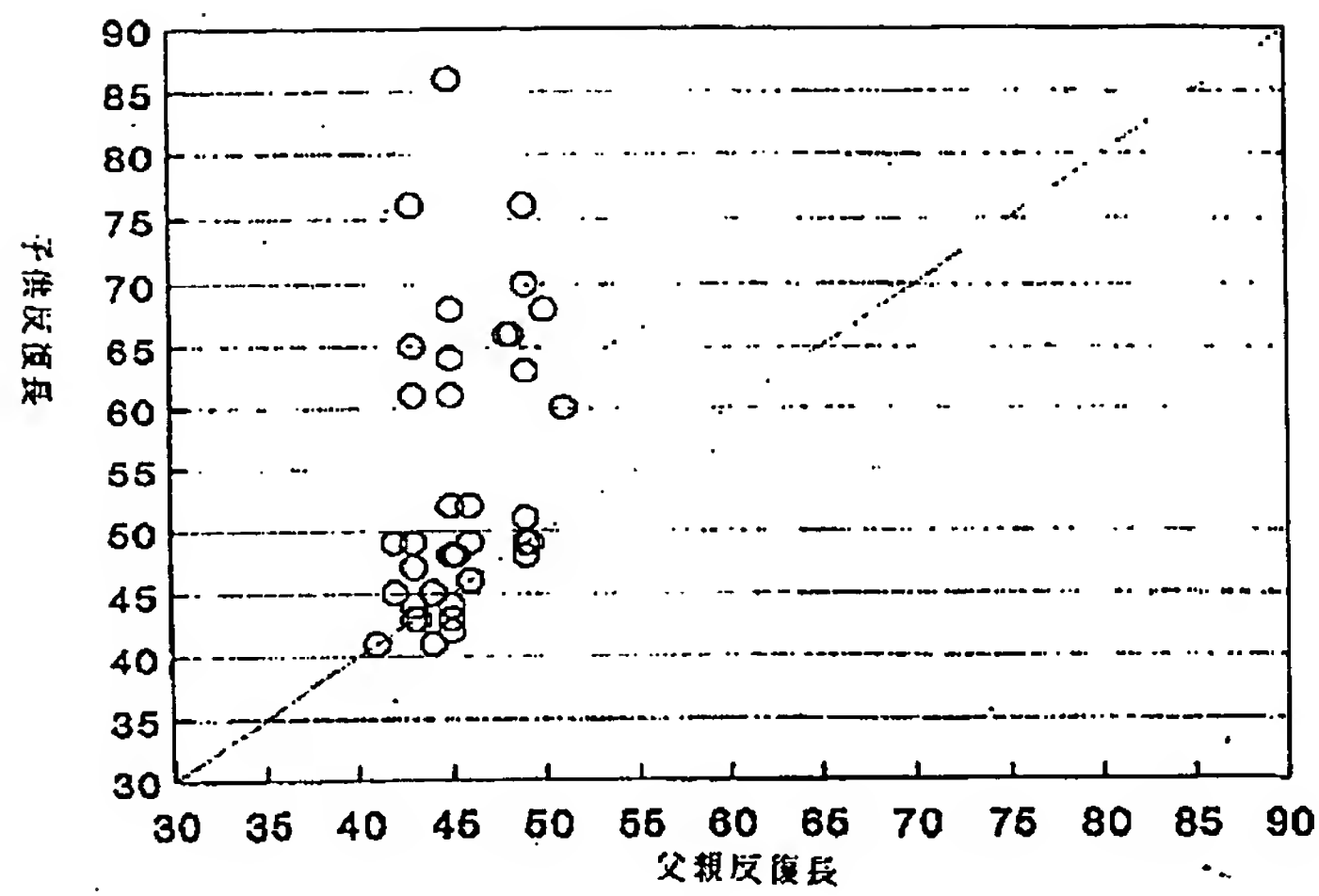
【図18】



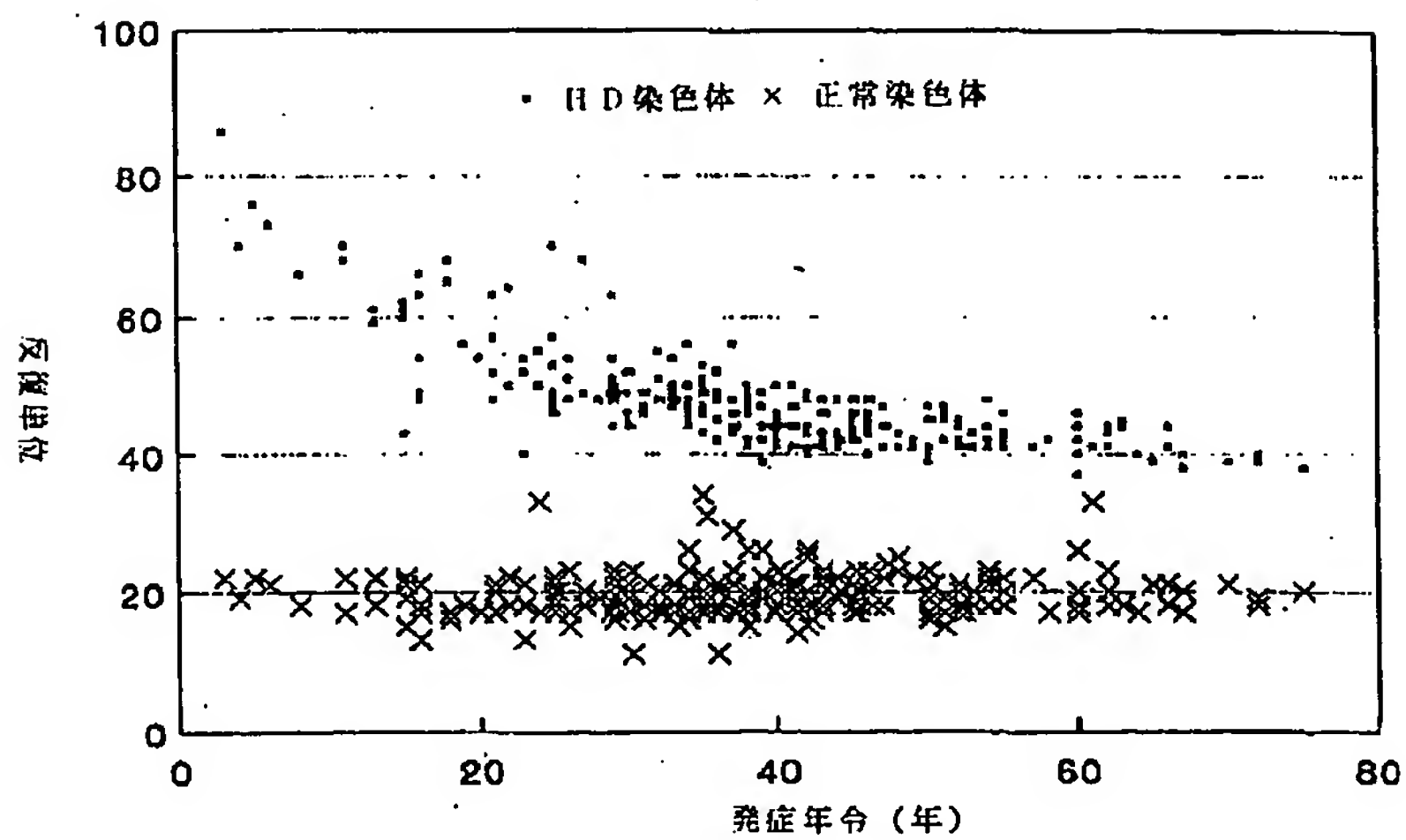
【図19】



【図20】



【図22】



フロントページの続き

(51)Int.Cl.<sup>6</sup>

C 07 K 16/18

C 12 P 21/02

21/08

C 12 Q 1/68

G 01 N 33/50

識別記号

庁内整理番号

F I

技術表示箇所

C 9282-4B

9161-4B

A 9453-4B

T

(72)発明者 クリスティーヌ・エム・アンブローズ  
アメリカ合衆国02129マサチューセッツ州  
チャールズタウン、エイトス・ストリート  
42番 ナンバー3105

(72)発明者 マーベル・ビー・ドゥヤオ  
アメリカ合衆国02138マサチューセッツ州  
ケンブリッジ、アバディーン・アベニュー  
24番

(72)発明者 ジェイムズ・エフ・ガセラ  
アメリカ合衆国01701マサチューセッツ州  
フレミングム、ウッドストック・ドライブ  
7番



THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-067661

(43)Date of publication of application : 14.03.1995

---

(51)Int.Cl. C12N 15/09  
A61K 38/00  
C07K 14/47  
C07K 16/18  
C12P 21/02  
C12P 21/08  
C12Q 1/68  
G01N 33/50

---

(21)Application number : 06-036026

(71)Applicant : GENERAL HOSPITAL CORP:THE

(22)Date of filing : 07.03.1994

(72)Inventor : MACDONALD MARCY E  
AMBROSE CHRISTINE M  
DUYAO MABEL P  
GUSELLA JAMES F

(30)Priority

Priority number : 93 27498  
93 85000Priority date : 05.03.1993  
01.07.1993

Priority country : US

US

---

(54) HUNTINGTIN DNA, ITS PROTEIN AND USE THEREOF

(57)Abstract:

PURPOSE: To obtain a new nucleic acid useful for gene therapy of Huntington's chorea or the like and for providing a recombinant DNA technique for diagnosis and treatment of Huntington's chorea,

CONSTITUTION: This isolated nucleic acid contains a nucleic acid encoding a huntingtin protein and is a gene spanning about 210 kb and encoding a protein having about 348 kDa. The gene is derived from the adjoining part of 500 kb segment between chromosome marker D4S180 and D4S182, is isolated from a cosmid contigs of the candidate region by a cloned trapped exon. A (CAG)<sub>n</sub> trinucleotide repeat varying from 37 to at least 86 copies is amplified by regulation, localization, stability or translatability of the mRNA.

---

## LEGAL STATUS

[Date of request for examination] 07.03.2001

[Date of sending the examiner's decision of rejection] 04.11.2003

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

THIS PAGE BLANK (374)

[Number of appeal against examiner's decision  
of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office



THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☒ **FADED TEXT OR DRAWING**

☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)